Chemical Engineering Science 280 (2023) 118972

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

A general procedure for the evaluation of the prediction fidelity of pharmaceutical systems models

Margherita Geremia [a], Samir Diab [b], Charalampos Christodoulou [c], Gabriele Bano [d], Massimiliano Barolo [a], Fabrizio Bezzo [a,*]

[a] CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, Via Marzolo 9, 35131 Padova, PD, Italy
[b] GlaxoSmithKline, Park Road, Ware SG12 0DP, UK
[c] GlaxoSmithKline, Gunnels Wood Road, Stevenage SG1 2NY, UK
[d] GlaxoSmithKline, 1250 S Collegeville Rd, Collegeville, PA 19426, USA

ARTICLE INFO

Keywords:
Systems model
Pharmaceutical development
Model identifiability
Parameter estimation
Model fidelity
Process engineering

ABSTRACT

Quantitative systems models have been increasingly used to accelerate the development of pharmaceutical processes that traditionally require time and resource-intensive experimental campaigns. However, despite the potential benefits of modelling in pharmaceutical process development, there is often a lack of confidence from stakeholders to adopt these approaches in a systematic way. This is particularly important in a highly regulated sector such as the pharmaceutical industry, where model uncertainty evaluation may have to be quantified and filed if model results are to be included as part of a regulatory submission. In this study, a systematic procedure is proposed, combining both standard techniques (e.g., global sensitivity analysis, model-based design of experiments) and new methods based on data analytics to assess model fidelity and to support practitioners in model usage for pharmaceutical development. A direct compression systems model for manufacturing oral solid dosage products is used as a case study. The systems model is comprised of the following sub-models: (1) Tablet press unit operation, (2) Tablet disintegration test unit, (3) In vitro dissolution test unit. The implementation of the methodology is critically discussed, showing the effectiveness of the approach and the benefits of the proposed techniques in *(i)* evaluating model fidelity as a result of uncertainty in parameter estimation, *(ii)* assessing whether it is necessary to estimate all systems model parameters in a statistically satisfactory way, and *(iii)* analyzing whether model calibration should be carried out on a unit basis or considering the overall system.

1. Introduction

Modeling in pharmaceutical manufacturing has been increasingly adopted to accelerate process development and optimization that typically require long and resource-intensive experimental campaigns (Destro and Barolo, 2022). Quantitative models can be used to represent the phenomena occurring along the manufacturing line and to examine the pharmacological effects of the drug product in the body, supporting rational drug design and decision making. The level of knowledge and understanding of the process, as well as the data availability required for calibration and validation of a model at a given stage of process development, can vary significantly, affecting the selection of the modeling method (Chen and Ierapetritou, 2020).

Pharmaceutical manufacturing processes are generally comprised of a set of unit operations connected by material and energy streams, which can be suitably represented through an integrated model (Boukouvala et al., 2012). Rather than exploiting the unit models separately for model-based research, an integrated model allows for investigating the influence of critical parameters in one unit to key performance indicators (KPIs) and critical quality attributes (CQAs) of the desired product downstream in the process line (Metta et al., 2019). The usage of mathematical models is not limited to the manufacturing process: mathematical models can describe test units assessing whether a pharmaceutical product meets the required specifications, or can be employed to represent the pharmacokinetic and pharmacodynamic behaviour of the active pharmaceutical ingredient (API) in the body. Such models, even if they are not physically connected to a unit operation and their inputs may not be material and energy streams derived from the manufacturing process, still maintain functional connections with the manufacturing model itself. For instance, these may comprise the transfer of physical properties and/or quality attributes. In this

* Corresponding author.
E-mail address: fabrizio.bezzo@unipd.it (F. Bezzo).

https://doi.org/10.1016/j.ces.2023.118972
Received 24 January 2023; Received in revised form 9 May 2023; Accepted 6 June 2023
Available online 10 June 2023

Nomenclature

Acronyms

API	active pharmaceutical ingredient
CI	confidence interval
CQA	critical quality attribute
DAE	differential and algebraic equations
DC	direct compression
DoE	design of experiments
GSA	global sensitivity analysis
HPLC	high performance liquid chromatography
IR	immediate release
KI	key indicator
KPI	key performance indicator
KS	knowledge space
LV	latent variable
MBDoE	model-based design of experiments
MC	Monte Carlo
OSD	oral solid dosage
PC	principal component
PCA	principal component analysis
PLS	partial least squares
R&D	research and development
RTD	residence time distribution
THz	terahertz
USP	Unites States Pharmacopeia
UV	ultraviolet

Greek letters

α	significance level
β	the total fraction of tensile strength that can be lost due to lubrication
γ	lubrication rate constant [dm^{-1}]
δ	Dirac delta function
ε	average porosity of the swollen product [−]
$\dot{\epsilon}$	erosion rate [m/s]
$\boldsymbol{\theta}$	set of parameters of the whole systems model
$\boldsymbol{\theta}_{M_i}$	set of parameters of sub-model M_i
$\boldsymbol{\Theta}$	regressor matrix of PCA/PLS models
λ	swelling rate [s^{-1}]
$\boldsymbol{\Lambda}$	diagonal matrix of the square roots of eigenvalue λ_a
λ_a	eigenvalue related to the a^{th} direction
μ	liquid viscosity [Pa s]
v_{ii}	ii^{th} term of the variance–covariance matrix
ρ_p	density of particles [kg m^{-3}]
$\boldsymbol{\Sigma}_i$	variance–covariance matrix of measurements errors in the i^{th} experiments
τ	total stress [MPa]
τ_{or}	average tablet tortuosity [−]
$\boldsymbol{\varphi}$	experiment design vector
ϕ	shape factor of particles [−]
ϕ^{ML}	maximum likelihood function
ψ	selected metric of $\boldsymbol{V}_\theta$ that represents the criterion for the experimental design
ω	lumped parameter for water penetration model [−]

Latin letters

A	number of principal components for PCA model
a_1	extended Kushner parameter (1)
a_2	extended Kushner parameter (2)
a_{sf}	Kawakita model parameter (1)
A_t	tablet surface area
B	number of latent variables for PLS model
b_1	extended Kushner parameter (3)
b_2	extended Kushner parameter (4)
B_{API}	rate of release of API
b_{sf}	Kawakita model parameter (2) [MPa^{-1}]
C_2	Peppas and Colombo parameter (1) [MPa]
C_3	Peppas and Colombo parameter (2) [MPa]
c_{API}	bulk concentration of API [kg m^{-3}]
c_{sat}	saturation concentration of API [kg m^{-3}]
d	length of the short axis of the tablet [m]
d_h	tablet hydraulic diameter [m]
E	elastic constant (1) [−]
$\boldsymbol{E}$	$\boldsymbol{\Theta}$-residuals for PCA model
$\boldsymbol{E_K}$	$\mathscr{K}$-residuals
$\boldsymbol{E}_0$	$\boldsymbol{\Theta}$-residuals for PLS model
$F_{B,N-B,0.95}$	95% percentile F-distribution with B, $N-B$ degrees of freedom
F_L	tablet hardness [MN]
G_0	elastic constant (2) [MPa]
H_{coat}	thickness of the coating layer [m]
K	extent of lubrication [dm]
k_{API}	mass transfer coefficient of API [$(m^3kg^{-1})^{nAPI}$ s^{-1}]
$\mathscr{K}$	response matrix of PLS model
$\boldsymbol{K}$	set of key indicators of the whole systems model
$\boldsymbol{K}_{m_j}$	set of key indicators of sub-model m_j
$\boldsymbol{K}^*$	vector of target key indicators
l	particle size [m]
$l_{0,API}$	particle size at the beginning of the process [m]
LC	percentage of label content [−]
m	number of sub-models of the systems model having associated at least one key indicator
M	number of sub-models of the systems model [−]
M_t	tablet mass [kg]
$M_{t,0}$	initial tablet mass [kg]
n	swelling parameter [−]
N	number of parameters combinations via MC simulations
n_{API}	order of dissolution [−]
N_{API}	number of dissolving API particles in particledistribution bin L and $L+dL$ size −
N_K	number of model key indicators
N_0	number of model parameters
N_{sp}	number of sampling intervals
N_y	number of measured variables
P	compaction pressure [MPa]
$\boldsymbol{P}$	$\boldsymbol{\Theta}$-loadings matrix for PCA model
$\boldsymbol{P}^{\boldsymbol{w}}$	weighed $\boldsymbol{\Theta}$-loadings matrix
p_c	capillary pressure [Pa]
P_d	water penetration depth [m]
$\boldsymbol{Q}$	$\mathscr{K}$-loading matrix
$\boldsymbol{R}$	$\boldsymbol{\Theta}$-loading matrix for PLS model
$R_{API,l}$	particle dissolution coefficient [m s^{-1}]
$\boldsymbol{S}$	$\boldsymbol{\Theta}$-score matrix for PLS model
s_b	semi-axis of the hyper-ellipsoid along the b^{th} direction
$s_{i,j}$	i,j^{th} element of the inverse of the measurement error covariance matrix
$S^{\lambda}_{i,j}$	similarity factor [−]
S_p	shape factor of pores [−]
sf	solid fraction [−]
t	time [s]
$\boldsymbol{t}$	solution on the score space
$t(\bullet)$	critical value of a t-distribution with $(1-\alpha/2)$ % confidence level and $(N-N_0)$ degrees of freedom
$\boldsymbol{t}_{new}$	direct inversion solution on the score space
$\boldsymbol{t}_{ref}$	reference t-value
t^{sp}	vector of the output variables sampling times
$\boldsymbol{T}$	$\boldsymbol{\Theta}$-scores matrix for PCA model
T^2_{lim}	95% confidence limit for Hotelling's statistics

T_t	tablet thickness [m]	w	wall height of the tablet [m]
$T_{t/2}$	half tablet thickness [m]	W	null space
TS	tensile strength [MPa]	$\boldsymbol{W}$	weight matrix
TS_0	tensile strength at zero porosity [MPa]	w_l	liquid content in the tablet [−]
$\boldsymbol{u}_{M_i}$	vector of input variables of sub-model M_i	x_{API}	mass fraction of API [−]
V_c	coating volume [m^3]	$\boldsymbol{x}_{M_i}$	vector of state variables of sub-model M_i
V_m	liquid volume in the vessel [m^3]	$\boldsymbol{y}$	vector of measured variables
$\boldsymbol{V}_\theta$	expected variance–covariance matrix of the model parameters θ	$\boldsymbol{y}_0$	vector of initial conditions for the measured variables
		$\widehat{\boldsymbol{y}}$	vector of model predictions
$\boldsymbol{V}_\theta{}^0$	preliminary parameter variance/covariance matrix	$\boldsymbol{y}_{M_i}$	vector of output measured variables of sub-model M_i

perspective, these models can be seen as additional sub-systems within a more comprehensive entity, which comprises, but is not limited to, the flowsheet model of the manufacturing process. The overall model comprising all sub-systems (e.g., all equations describing the relevant phenomena occurring in a single unit operation, or in a single in vitro product performance test unit) is often referred to as a *systems model* (Avraam et al., 1998).

Over the last 15 years, several examples of pharmaceutical manufacturing systems model applications have been proposed in the literature. Just to mention some recent contributions, Bano et al. (2022) streamlined the development of an industrial dry granulation process for immediate release tablets. White et al. (2022) developed a systems model of a pharmaceutical tablet manufacturing process comparing dry granulation with direct compression. Moreno-Benito et al. (2022) presented a hybrid systems model of a continuous direct compression line combining first-principles and data-driven approaches. In Yang et al. (2022) a systems model is exploited to design a recombinant adeno-associated virus drug manufacturing process operating in batch and continuous mode. Diab et al. (2022a) proposed a general framework to guide control strategies for continuous manufacturing of an API in an industrial environment. In Diab et al. (2022b), an integrated approach was used to find the optimal process setpoints to optimize the manufacturing process of API production. Destro et al. (2022) proposed a benchmark simulator for the development and testing of quality-by-design and quality-by-control strategies in continuous integrated filtration-drying of crystallization slurries.

Despite the potential benefits of modeling in pharmaceutical process development and decision making, the systematic use of these methods is not as widespread as it could be. It is not uncommon that stakeholders are apprehensive towards the adoption of quantitative models, and that is mainly due to the lack of confidence in the prediction capability with respect to KPIs and CQAs (Braakman et al., 2022). Another aspect is that pharmaceutical processes typically involve very complex phenomena that may not be easily captured by first-principles models – this is particularly true for the manufacturing of small molecules (e.g., in multiphase reactions with heterogeneous catalysts) and biopharmaceutical products. Therefore, resorting to empirical models built on design of experiments (DoE) is deemed to be less risky. However, high in-house expertise and sufficient resources to develop the first-principles model in question may be needed.

The application of standardized model evaluation methods (i.e., quantitative and statistical criteria for the assessment of model prediction uncertainty) is required to enhance the adoption of quantitative models for pharmaceutical manufacturing development and decision making (Zineh, 2019). The aim is ensuring a pre-set confidence in the prediction of the model KPIs and CQAs (that from now on will be generically named key indicators (KIs)). This is fundamental in a sector that is as highly regulated as the pharmaceutical industry, and can be used to support the submission of future regulatory filings for new assets. The identification of a rigorous workflow to assess the performance of quantitative models has been explicitly identified as a central challenge to be addressed (Bai et al., 2019). This is particularly relevant to process systems models that are being developed in the context of a pharmaceutical industrial environment where resources may be limited for the model development, but high prediction confidence is required: in order to frame methods and techniques available in the literature, Braakman et al. (2022) recently proposed a concept paper discussing the evaluation of systems models in general terms.

Considering how expensive pharmaceutical research and development (R&D) is, and the timelines in which one must develop and launch a product due to its limited patent life (Destro and Barolo, 2022), it is essential that a model be developed efficiently as to maximize the benefits for process development. This means: (*i*) developing the model quickly, (*ii*) minimizing the resources required for its development (in terms of material and labour), and (*iii*) ensuring that the confidence in the model predictions is sufficient for its intended purpose and impact. Whenever a reliable mathematical model is available, the prediction fidelity with respect to the KIs (which are typically a subset of all model outputs) is strongly affected by the precision of model parameter estimates. In a systems model, not only does the prediction fidelity of a selected KI depend on the parameters of the specific sub-model, but it also depends on the parameters related to the previous sub-models impacting the KI of interest. For instance, let us consider the manufacturing of a small molecule API, which is synthesized in a reactor (where some impurities are also formed), undergoes some purification (e.g., by liquid–liquid extraction) and finally is isolated and separated as a solid in a crystallizer. The prediction of the level of critical impurities (that may be genotoxic or mutagenic, and thus are key components of the quality target product profile) in the drug substance will be impacted by the amount of these impurities formed in the upstream reactor. Therefore, it is expected that uncertainty in the parameters of the reactor model will impact the predicted levels of impurity in the isolated drug substance (Diab et al., 2022a).

In this context, the following questions need to be addressed:

(i) Since accounting for uncertainty in model predictions is fundamental, how can we quantitatively assess the model fidelity, so as to ensure an assigned confidence in the prediction of the model KIs?
(ii) Should all model parameters be estimated in a statistically satisfactory way, or is it sufficient to focus on just a subset of them? What is the parametric precision that must be attained to satisfy the prediction requirements for the KIs?
(iii) Integrated models are currently calibrated on a unit-by-unit basis, i.e., fit-for-purpose data generation for each unit sequentially. Is this really the most efficient strategy?

In this study, we propose a model evaluation framework to answer the questions above and to support the critical usage of quantitative models within a pharmaceutical manufacturing environment. The framework aims to:

(i) Assess model prediction fidelity using standardized model evaluation methods, i.e., to quantify the impact of model parameter uncertainty on the selected model KIs;

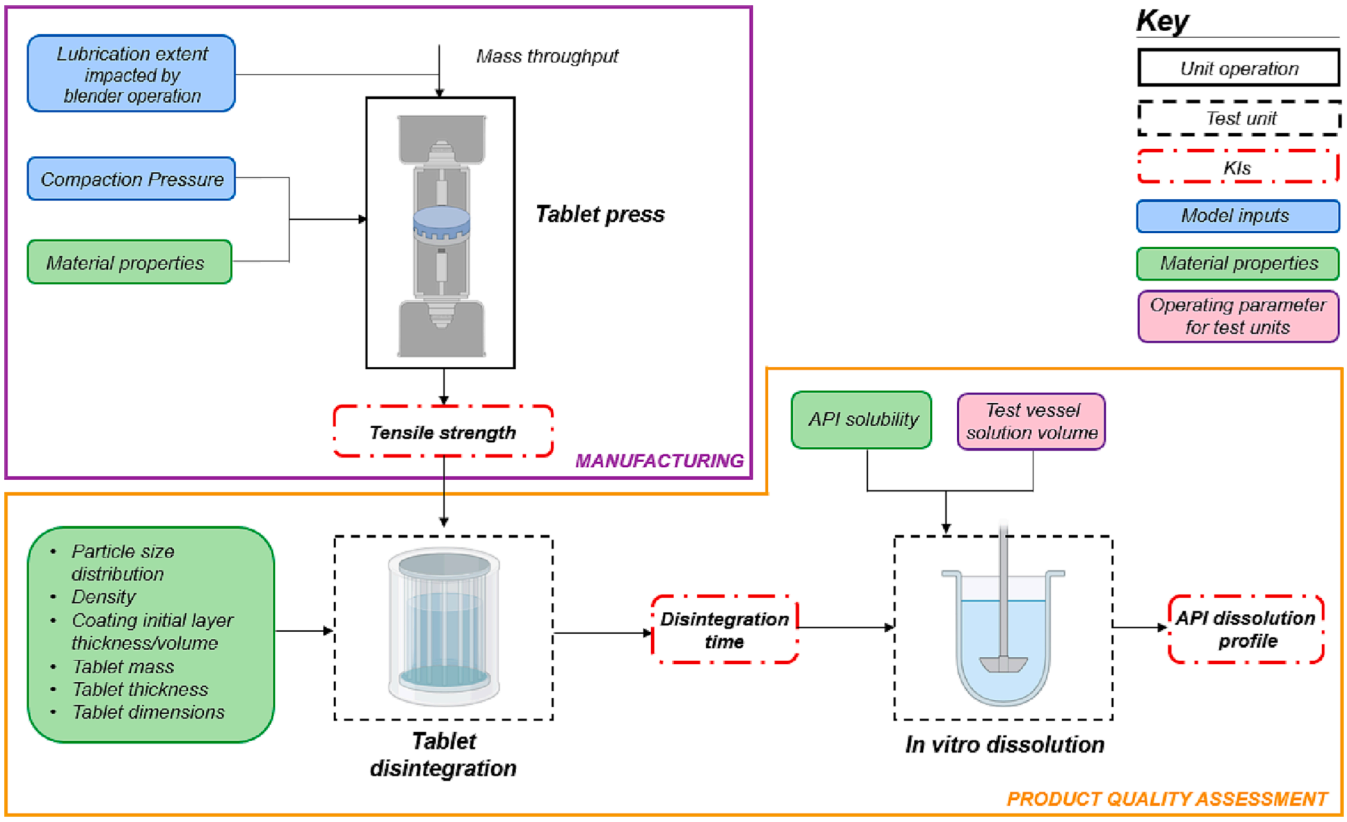

Fig. 1. Integrated entities of the direct compression systems model considered in this study: Tablet press unit operation, Tablet disintegration test unit, and In vitro dissolution test unit. The KIs for each sub-system are reported within red dashed boxes, i.e., tensile strength, disintegration time, and API dissolution profile. Black solid boxes are unit operations; black dashed boxes are test units; blue boxes identify model inputs; green boxes identify material properties; violet boxes identify operating parameters for test units.

(ii) Plan the experimental campaigns for the parameter estimation in quantitative (systems) models, i.e., optimize the experimental campaigns to maximize the information with the minimal experimental effort.

We combine techniques that are typically adopted in different modeling contexts – namely, model-based design of experiments (MBDoE) for parameter estimation, and data reduction by means of multivariate statistical methods to enhance the interpretation of the results. A direct compression (DC) systems model for manufacturing oral solid dosage (OSD) products will be used as a case study. The systems model is comprised of the following sub-models: (1) Tablet press unit operation, (2) Tablet disintegration test unit, and (3) In vitro dissolution test unit. The implementation of the methodology is critically discussed, showing the effectiveness and the benefits of the proposed techniques.

The article is organized as follows. In Section 2, we introduce the framework for the assessment of model prediction fidelity, and we briefly outline the general scope of each step of the procedure. In Section 3, we thoroughly describe the suggested techniques to implement the methodology, and we present two possible workflows to be used for the evaluation of integrated models. In Section 4, we give details about the DC systems model for manufacturing OSD products that is used as the case study. In Section 5, we implement the proposed methodology for the evaluation of the prediction fidelity of the DC systems model, and we critically discuss the results that are obtained. Some final remarks will conclude the study.

2. Framework for model evaluation

Let us consider the systems model represented in Fig. 1, where a DC system for manufacturing OSD products is illustrated; the same systems model will be used as a case study in this work. Here, we are not considering blend or content uniformity or tablet weight variability, and thus the only sub-models we consider are (1) the tablet press unit operation, (2) the tablet disintegration test unit, and (3) the in vitro dissolution test unit. Note that the tablet press model is the only sub-system concerning a unit operation in the manufacturing process; however, the methodology we propose is general and thus the results are not affected by omission of the other unit operations in a typical DC line (i.e., feeder hoppers, screw feeders, blenders, transfer hoppers, tablet coater). The other sub-systems represent experimental tests for the assessment of product CQAs, and require information from the tablet press model, i.e., the lubrication extent attained in the upstream powder blending, and the compaction pressure exerted by the press. In this example, each sub-system outputs a KI, i.e., the tensile strength from the tablet press unit operation, the disintegration time from the tablet disintegration test unit, and the API dissolution profile from the in vitro dissolution test unit.

2.1. General procedure

In mathematical terms, a generic systems model is comprised of a number M of models representing the different sub-systems (M_1, M_2, …M_M), and the relationship between the unit inputs and outputs can be described by a set of differential and algebraic equations (DAEs):

$$\boldsymbol{f}_{M_i}\left(\boldsymbol{x}_{M_i}(t), \dot{\boldsymbol{x}}_{M_i}(t), \boldsymbol{\theta}_{M_i}, \boldsymbol{u}_{M_i}(t), t\right) = 0,$$

$$\boldsymbol{y}_{M_i} = \boldsymbol{g}_{M_i}(\boldsymbol{x}_{M_i}(t)), \tag{1}$$

$$\boldsymbol{K}_{m_j} = \boldsymbol{h}_{m_j}(\boldsymbol{x}_{m_j}(t)) \text{ with } i = 1, 2, \cdots, M; j = 1, 2, \cdots, m \text{ where } \boldsymbol{x}_{M_i},\ \boldsymbol{u}_{M_i},\ \boldsymbol{y}_{M_i}$$

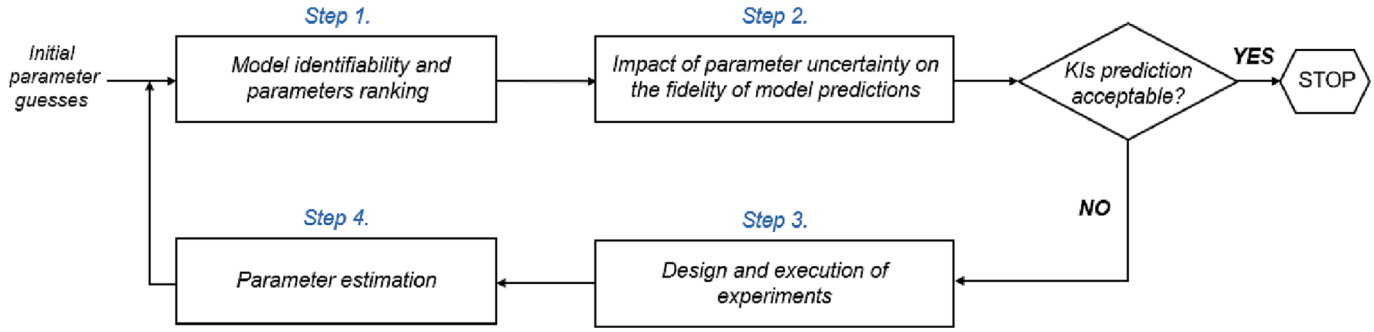

Fig. 2. Schematic of the proposed general framework to quantify the model parameter impact on the prediction fidelity of model KIs.

and $\boldsymbol{\theta}_{M_i}$ refer to sub-model i and are (respectively) the vector of state variables, the vector of input variables, the vector of measured output variables, and the vector of the model parameters. We indicate with $m \leq M$ the number of sub-models having an associated vector of KIs, and $\boldsymbol{K}_{m_j}$ the vector of KIs for the sub-model j. Here, we assume that the vector $\boldsymbol{u}_{M_i}$ of input variables coincide with the control variables, i.e., the variables we can manipulate to control the KIs. A KI can be equal to a measured output y_{t,M_i} or be derived from combinations of outputs/inputs. With reference to the entire systems model, we will denote the set of all model parameters θ and model key indicators $\boldsymbol{K}$ as follows:

$$\boldsymbol{\theta} = \left[\boldsymbol{\theta}_{M_1}, \boldsymbol{\theta}_{M_2}, \cdots, \boldsymbol{\theta}_{M_M}\right]^T \tag{2}$$

$$\boldsymbol{K} = \left[\boldsymbol{K}_{m_1}, \boldsymbol{K}_{m_2}, \cdots, \boldsymbol{K}_{m_m}\right]^T \tag{3}$$

Assessing the fidelity of a given systems model requires quantifying how the contributions of all model parameters impact the prediction of the KIs. In general, the prediction fidelity depends not only on the parameters of the specific sub-system model, but also on the parameters of all sub-system models impacting the unit being investigated.

The methodology consists of four sequential steps (Fig. 2) that are iteratively repeated until satisfaction of the stop criterion, which is attained when all predictions of model KIs are within the desired tolerance.

Step 1. Model identifiability and parameters ranking.

Before any parameter estimation task can be performed, an identifiability analysis should be performed to determine whether it is possible to formally estimate the systems model parameters (Miao et al., 2011). Once it has been verified that all parameters can be estimated, the objective is assessing their impact on the KIs predictions. In other words, it is necessary to *rank* the importance of parameters with respect to each KI of interest in such a way as to prioritize the subsequent experimental effort (Step 3) for parameter estimation towards the most influential parameters (Saltelli et al., 2007).

Step 2. Impact of parameter uncertainty on the fidelity of model prediction.

The second step aims at *quantifying* how uncertainty in parameter values impacts the prediction fidelity of the KIs: given the current uncertainty on estimated values of model parameters, the goal is to assess whether the attained precision on the KIs prediction is acceptable or not. The stop criterion is satisfied when all model predictions of the KIs fall within the range of desired tolerance, and therefore no further improvement in the precision of estimated parameters is needed.

Step 3. Design and execution of experiments.

If the parameter precision needs improving (i.e., stop criterion not satisfied at Step 2), additional experiments should be designed, so that model parameters can be estimated on the basis of the new experimental observations. Experiments are planned to facilitate the parameter estimation task – namely, to identify all the model parameters with the minimal experimental effort. To this purpose, the design of a new experiment should follow a rational criterion, i.e., experimental conditions should be selected to generate the most informative data for model parameterization (Bard, 1974). The new experiment is executed using the experimental conditions planned by design.

Step 4. Parameter estimation.

Once new data are available, model parameters are re-estimated. Most estimation methods require to find values of model parameters for which some objective function attains its maximum or minimum (Bard, 1974). After the estimation procedure, the reliability and statistical precision of estimates should be investigated; statistical hypothesis testing methods have been extensively discussed by Anderson (1958).

2.2. Workflow

Different workflows are possible in order to apply the procedure presented in this study. We implement and comment on two of them, namely (*i*) a *modular* approach and (*ii*) a *global* approach.

In a modular approach, the KIs of all units are targeted sequentially. Model parameters are estimated on a sub-system basis. Practically, we aim at obtaining a reliable prediction of KIs following the process layout. For instance, let us consider the process in Fig. 1. First, we focus on the tablet press unit operation, where the KI of interest is the tablet tensile strength (i.e., hardness). Once the tablet press model parameters are precise enough to guarantee sufficient tensile strength prediction fidelity, we move on to the tablet disintegration test unit. When we reach the required precision of the relevant parameters, we finally focus on the API dissolution profile from the in vitro dissolution test unit. A modular approach can be convenient to organize the experimental campaign, since experiments are typically designed and implemented for one unit at a time. On the other hand, it is assumed that the parametric precision attained on that unit will be sufficient to reach the KI fidelity required in subsequent units. If this is not the case, an iterative experimental procedure may be required.

In the global approach, all KIs are targeted simultaneously, and the parameters of all sub-system models impacting the KIs are considered at the same time. For the process in Fig. 1, this means that we look at the tablet press unit operation, the tablet disintegration test unit, and the in vitro dissolution test unit simultaneously. The estimation and design of experiments tasks are performed considering all relevant parameters in the systems model, until the desired fidelity for all model KIs (i.e., tensile strength, disintegration time, and API dissolution profile) is achieved. This approach is potentially more efficient than the modular one as no iteration over different sub-systems is needed, and the impact of all parameters on KIs is comprehensively assessed. However, the implementation of the experimental campaign may be more difficult as different experiments for different units need to be performed simultaneously: for example, for a given iteration of the workflow, tablets are made in the press, then run in the disintegration test, and finally their dissolution profile is measured.

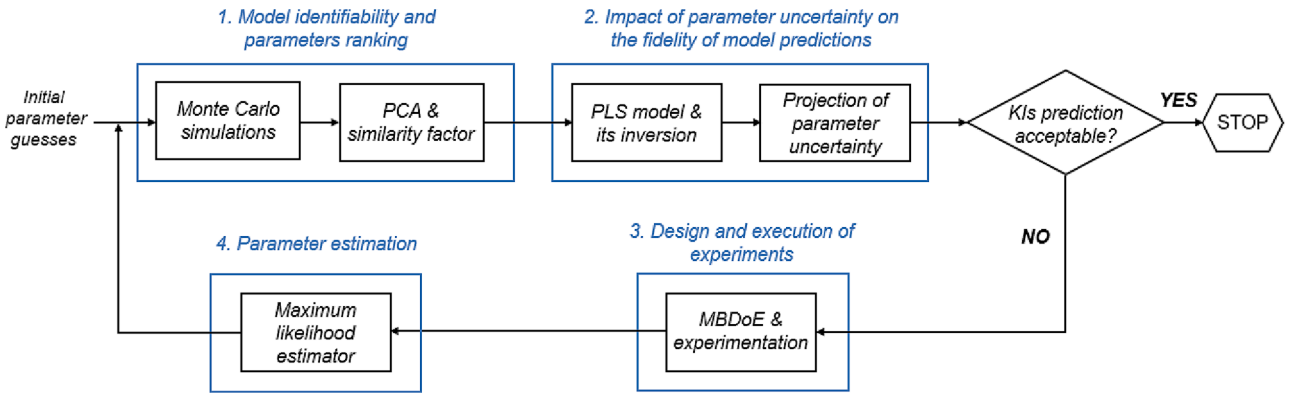

Fig. 3. Schematic of suggested techniques to accomplish steps 1–4 of the proposed general framework.

3. Implementation of the methodology

In this section, we present the suggested techniques to implement steps 1–4 of the general framework discussed in Section 2. We combine methods used in different modeling contexts – namely, MBDoE, and data reduction by means of multivariate statistical methods (i.e., principal component analysis (PCA), and partial least-squares regression (PLS)), as shown in Fig. 3.

3.1. Model identifiability and parameters ranking

A discussion on structural model identifiability is beyond the scope of this work, and we refer to Miao et al. (2011) and Braakman et al. (2022) for a thorough analysis. However, we would like to point out that in many practical cases, the assessment of the model structural (or global) identifiability, i.e., the theoretical possibility to estimate all model parameters, can be difficult to prove and one often needs to rely on sensitivity techniques, which are discussed in the following. Furthermore, global identifiability may be unnecessary, because the goal is to assess model fidelity with respect to some predefined KIs. For instance, let us suppose that a parameter cannot be estimated from available measurements (i.e., the model is not identifiable), but that parameter has little or no impact on the prediction of a KI of interest: in such a case, its estimation is of no practical relevance.

Different techniques can be used to evaluate the influence of parameter estimates towards the prediction of selected KIs. Estimability analysis (McLean and McAuley, 2012) or, most commonly, local and global sensitivity analyses, are the practical tools to assess a model's (local) identifiability. Sensitivity analyses provide metrics to rank the importance of parameters with respect to an output of interest: the larger the values of the metrics, the more sensitive the model response with respect to parameter changes. Local sensitivity analysis focuses on the sensitivity of the outputs to the perturbation of one single parameter (at a time) around a particular value. When very little preliminary information on parameter values is available, performing local sensitivity analysis is not recommended (Saltelli et al., 2007). On the other hand, global sensitivity analysis (GSA) measures the uncertainty in the output caused by the uncertainty in the parameter values over a range of possible values. GSA methods can be classified as derivative-based (e.g., Morris' (1991) method), correlation-based (e.g., partial rank correlation coefficient method, (Iman and Helton, 1988)), and variance-based (e.g., Fourier amplitude sensitivity test, (Cukier et al., 1973), Sobol's (1993) method). Variance-based sensitivity analyses are frequently used; they decompose the variance in the model output and allocate it to each model parameter. Two widely used metrics are the first-order and total effects (Homma and Saltelli, 1996). In Section S.1 of the Supplementary material, a short description of such metrics and the results of their implementation in our case study are shown.

One limitation of the classical variance based Sobol's GSA is that it is defined under the assumption of independent parameters, which might lead to unreliable results if correlations are present (Song et al., 2016), unless more sophisticated GSA techniques (e.g., Xie et al., 2019; Barr and Rabitz, 2023) are used. In the following, we propose an alternative approach based on PCA, which allows (*i*) simplifying the description and graphical representation of parameter combinations, and (*ii*) visualizing their correlation structure.

3.1.1. PCA analysis for parameter ranking

The goal is to assess how the fidelity of one KI of interest depends on the N_θ model parameters that are affected by uncertainty. The following approach is proposed:

(i) given the initial parametric set θ, we determine the control variables $\mathbf{u}$ in such a way that the predicted KI based on the current parameter values is equal to the target value, and we fix them;
(ii) once vector $\mathbf{u}$ has been fixed, the uncertainty of the predicted KI only depends on the uncertainty of model parameter estimates; given the parameter range of variability, we account for their uncertainty by generating a sufficiently high number (Kucherenko et al., 2015) of parameter combinations using a Monte Carlo (MC) method, with probabilistically selected model parameters (Fishman, 1995). A uniform distribution of the parameter values is assumed;
(iii) in order to evaluate the influence of model parameters on KI prediction, the original (sub-system or systems) model is directly exploited to compute the values of the KI for the correspondent parameter combinations. We retain the set of N parameter combinations, for which the correspondent KI prediction falls within $\pm$ 50% the target value.
(iv) At this point, PCA (Montgomery, 2013) is used to summarize the information of the matrix $\boldsymbol{\Theta}$ $[N \times N_\theta]$ of the N combinations of N_θ model parameters by projecting them onto a new coordinate system of independent variables called principal components (PCs), aiding visualization and graphical interpretation.

To avoid the scaling effect of different orders of magnitude of the model parameters, $\boldsymbol{\Theta}$ is autoscaled, i.e., data are mean-centered and scaled to unit variance (Wise and Gallagher, 2006). The representation of $\boldsymbol{\Theta}$ in PCA is given by:

$$\boldsymbol{\Theta} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E}, \tag{4}$$

where $\boldsymbol{T}$ $[N \times N_\theta]$ is the matrix of scores, $\boldsymbol{P}$ $[N_\theta \times A]$ is the matrix of loadings, and $\boldsymbol{E}$ $[N \times N_\theta]$ is the matrix of residuals.

The model parameters can be plotted as points using their loadings on PCs as coordinates. If some parameters in $\boldsymbol{\Theta}$ are correlated, one common direction of variability can be identified, which is described through a single PC. Therefore, we can represent the variance of $\boldsymbol{\Theta}$ using

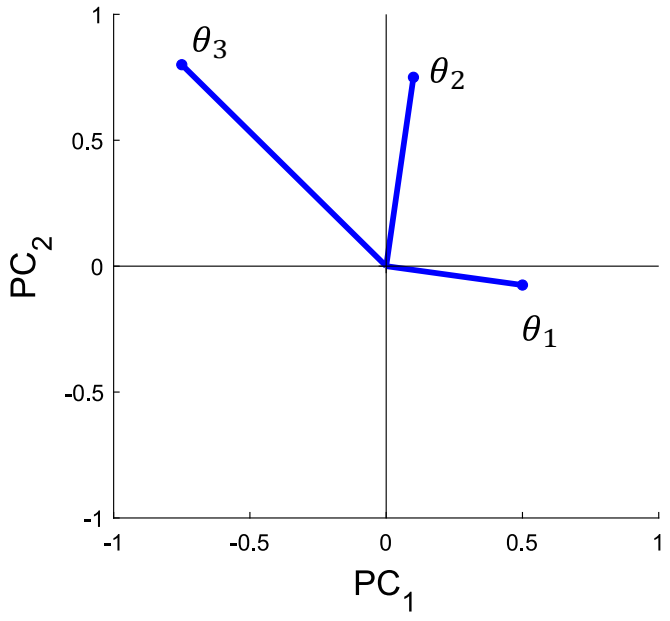

Fig. 4. Illustrative example of PCA application for parameters ranking, where three model parameters are projected onto a coordinate system of two PCs. The identification of the most influential contribution (parameter θ_3 is based on the length of parameter loadings (blue segments), while the correlation structure is described by their mutual directions.

a number A of PCs, which is smaller than the number of parameters, N_θ. In this work, we select A according to the "eigenvalue greater than one" rule (Kaiser, 1991). Other criteria can be found in the literature (Zwick and Velicer, 1986). The Mahalanobis distance (Mahalanobis, 1930) is used to measure their distance from the origin of the coordinate system of PCs.

As an example, let us consider the simple case represented in Fig. 4, where three parameters are projected onto a coordinate system of two PCs. For the sake of simplicity, we can assume that $A = 2$ is sufficient to capture enough variance of data in $\boldsymbol{\Theta}$, and that PC_1 and PC_2 capture an equal percentage of variance (the axes scales in Fig. 4 are equal). It is evident that parameters θ_1 and θ_2 are not correlated (orthogonal loadings); parameter θ_3 equally contributes to PC_1 and PC_2. Some correlation between pairs θ_1-θ_3 and θ_2-θ_3 can be observed, and this may lead to potential difficulties in their estimation. The uncertainty in the prediction of the KI of interest can be allocated to each model parameter by identifying the most influential contributions according to the length of the corresponding projected loading along the PCs; parameter θ_3 can be identified as the most impactful.

3.1.2. Use of similarity factor to assess the influence of operating conditions

The proposed methodology assumes that control variables **u**, i.e. the experimental settings and/or operating conditions in the manufacturing process, are fixed. This is quite an important assumption, which is common to all methodologies, including sensitivity analyses: a (subsystem or systems) model is always assessed at a given operational status. In our framework, given a parametric set **θ**, control variables **u** are fixed in order to attain the target values of the KIs.

However, along subsequent iterations of the procedure in Fig. 3, the estimated values of the parameters may change, and therefore **u** should adapt accordingly, providing new trajectories for time-variant control variables and new settings for time-invariant ones. In other words, **u** should adjust according to the model parameterization. The consequence is that, in general, the correlation structure of model parameters depends on control variables. Stated differently, this means that the relative importance of parameters (and their correlation) cannot be separated by process operation. This is not surprising in model that is nonlinear in its parameters. As a matter of example, consider a reactive system whose reactions kinetics are described using the standard Arrhenius expression: note that the importance of activation energies with respect to other parameters (e.g., the pre-exponential factors) increases when temperature increases.

One advantage of the PCA methodology described in Section 3.1.1 is that we can exploit established techniques to analyze how the correlation structure of model parameters varies when operating conditions change; namely, we can assess how variations in control variables affect the interaction between parameters and their relative ranking. In particular, different PCA models for matrices $\boldsymbol{\Theta}_i$ $[N_i \times N_\theta]$ and $\boldsymbol{\Theta}_j$ $[N_j \times N_\theta]$ (i.e., obtained using two sets, i and j, of different fixed operating conditions) can be compared through the similarity factor $S^{\lambda}_{i,j}$ (Gunther et al., 2009):

$$S^{\lambda}_{i,j} = \frac{trace\left[\left(\boldsymbol{P}^{w}_{i}\right)^{T}\left(\boldsymbol{P}^{w}_{j}\right)\left(\boldsymbol{P}^{w}_{j}\right)^{T}\left(\boldsymbol{P}^{w}_{i}\right)\right]}{\sum_{a=1}^{A}\lambda_{i,a}\lambda_{j,a}}, \tag{5}$$

where $\boldsymbol{P}^{w}_{i}$ and $\boldsymbol{P}^{w}_{j}$ are the corresponding weighed loading matrices. $\boldsymbol{P}^{w}_{i}$ is defined as:

$$\boldsymbol{P}^{w}_{i} = \boldsymbol{P}_i\boldsymbol{\Lambda}_i, \tag{6}$$

with $\boldsymbol{\Lambda}_i$ $[A \times A]$ being the diagonal matrix of the square roots of the eigenvalues $\lambda_{i,a}$.

$S^{\lambda}_{i,j}$ is always bounded between 0 and 1. When $S^{\lambda}_{i,j} = 0$, there is no similarity between the two PCA models, which means that when settings and operating conditions change, the parameter correlation structure changes, too, and therefore the analysis on parameter ranking should be performed again. Conversely, when $S^{\lambda}_{i,j}$ is close to 1 there is high similarity between the two PCA models, thus implying that the correlation structure of model parameters is scarcely dependent on the operating conditions. In Section S.2 of the Supplementary material, a case, where process conditions affect the correlation between parameters, are discussed as a matter of example.

3.2. Impact of parameter uncertainty on the fidelity of model predictions

The goal is quantifying the impact of parameter uncertainty on the fidelity of model predictions on KIs. Given the current parameter uncertainty together with their estimated values, the original model can be directly used to express the relationship between model parameters and KIs, while accounting for the uncertainty of model predictions through a Monte Carlo (MC) method (Fishman, 1995). It is worth noting that MC simulations require the generation of a sufficient number of model parameters combinations that typically are pseudo-random sequences or quasi-MC with Sobol sequences (Kucherenko et al., 2015). Recently, Montes et al. (2018) proposed a modeling activity coupled with MC simulations to improve the design and optimization for the production of ibuprofen, while Tian et al. (2022) characterized the uncertainty in the prediction of the residence time distribution (RTD) for a continuous powder blending process via MC sampling of the RTD parameters.

The above methodology can in principle be applied for the purpose of this work, too. However, to improve interpretability of results, here we propose an alternative approach. Our methodology is based on a PLS regression model (Wold et al., 1983; Geladi and Kowalski, 1996) relating the model parameters and the KIs of interest.

Before introducing the method formally, let us explain the scope qualitatively. If the model parameters were estimated perfectly, the only issue would be to manipulate the control variables so that the KIs are attained. A process operated in that way would target the KIs. However, there is always uncertainty in the estimated values of parameters. Thus, the question is: considering the range of variability determined by the precision of parameter estimation, can we guarantee that KIs will be predicted within a pre-defined acceptable tolerance? The procedure

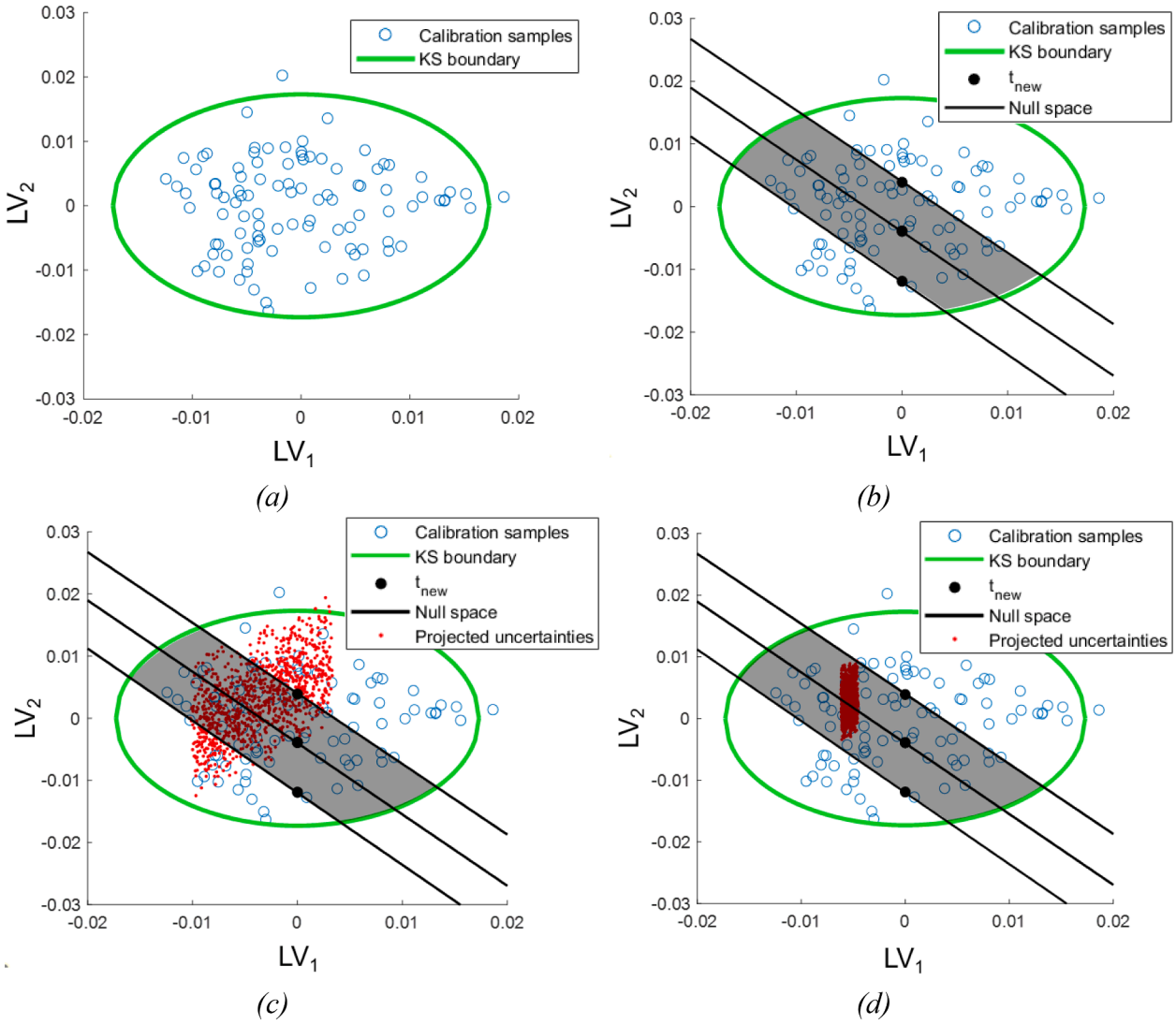

Fig. 5. Illustrative example of PLS model for the assessments of the impact of parameter uncertainty on the fidelity of model predictions. (a) Projection of the calibration samples (light blue circles) onto the score space and KS boundary (green ellipse). (b) Inversion solutions t_{new} (black points) and null spaces across them (black lines) for the inequality constraints problem, which define the KI acceptability region (grey area). (c) Case in which some projections of parameter uncertainty (red points) fall outside the KI acceptability region. (d) Case in which all projections of parameter uncertainty (red points) fall inside the KI acceptability region.

discussed in the following will aim at determining the parametric space that for given control variables **u**, allows predicting the KIs with the required fidelity. The procedure will determine the space containing all parametric sets guaranteeing that if the process is carried out in a certain way (set by **u**), then the KIs will be targeted. Conversely, if the uncertainty in the parameter estimates is such that the actual values of the parameters may lay outside the acceptability region, then a higher precision will be needed.

Let $\boldsymbol{\Theta}$ $[N \times N_\theta]$ be the same input (regressor) matrix of N combinations of N_θ model parameters we have defined in Section 3.1 and generated using MC simulations – where a sufficiently high number of scenarios was selected according to Kucherenko et al. (2015), and a uniform distribution of parameters values was assumed. Note that if more information is available on the actual distribution of parameter uncertainties, more rigorous approaches should be used (Schenkendorf et al., 2018). $\mathscr{K}$ $[N \times N_K]$ is the correspondent response matrix of N combinations of N_K KIs, that are computed using model (1).

PLS is a multivariate regression technique that projects the regressor and response variables onto a common latent space, according to the model structure:

$$\boldsymbol{\Theta} = \boldsymbol{S}\boldsymbol{R}^T + \boldsymbol{E}_\theta, \tag{7}$$

$$\mathscr{K} = \boldsymbol{S}\boldsymbol{Q}^T + \boldsymbol{E}_K, \tag{8}$$

$$\boldsymbol{S} = \boldsymbol{\Theta}\boldsymbol{W}. \tag{9}$$

$\boldsymbol{S}[N \times N_\theta]$ is the score matrix, $\boldsymbol{R}$ $[N_\theta \times B]$ and $\boldsymbol{Q}$ $[N_K \times B]$ are matrices of loadings, while $\boldsymbol{E}_\theta$ and $\boldsymbol{E}_K$ the residuals. $\boldsymbol{W}$ $[N \times B]$ is the weight matrix, through which the data in $\boldsymbol{\Theta}$ are projected onto the latent space to give $\boldsymbol{S}$. B is the number of significant latent variables (LVs) chosen to build the model; namely, it corresponds to the dimension of the model space. B must be selected as to capture most of the variance of the input and output data, and is usually chosen through a cross-validation procedure (Wold et al., 2001). Confidence limits are considered in the latent space in the shape of a hyper-ellipsoid, which defines the knowledge space (KS) boundary (MacGregor and Bruwer, 2008); inference using the PLS model can be made only for points that lie inside this region. The KS boundary is mathematically described by the 95% confidence limit of the Hotelling's T^2 statistic (Hotelling, 1993):

$$T^2_{lim} = \frac{B(N-1)}{N-B} F_{B,N-B,0.95}, \tag{10}$$

where $F_{B,N-B,0.95}$ is the 95% percentile of the F-distribution with B and $(N-B)$ degrees of freedom.

The semi-axis s_b of the hyper-ellipsoid along the b^{th} direction can be described as:

$$s_b = \sqrt{\lambda_b T^2_{lim}} \tag{11}$$

where λ_b is the eigenvalue of the matrix $\boldsymbol{\Theta}^T \mathscr{K} \mathscr{K}^T \boldsymbol{\Theta}$ related to the b th direction.

Once again, it should be stressed that control variables **u** are fixed in order to target the KI values based on the current parametric set. This implies that the fidelity of model predictions only depends on the uncertainty of the parameter estimates. The PLS model can be inverted (Jaeckle and MacGregor, 2000) to determine the set of parameters combinations that guarantee that the prediction of the KIs of interest is within the range of desired tolerance. The PLS model inversion provides an immediate graphical reading of results, which can efficiently support numerical interpretation. The quality target is described by inequality constraints, i.e., the bound values for the response prediction, which in turn define the KIs acceptability region. If the dimension of the latent space B is equal to N_{KI}, one unique solution $\boldsymbol{t}_{new}$ for the target response variable of KIs, $\boldsymbol{K}^*$, exists:

$$\boldsymbol{t}_{new} = \left(\boldsymbol{Q}^T\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}^T\boldsymbol{K}^{*T} \tag{12}$$

When the dimension of the latent space B is greater than N_K, we obtain multiple solutions of the model inversion problem, which are defined by:

$$W = \{(\boldsymbol{t}_{new} + \boldsymbol{t}), \boldsymbol{t} \in \ker(\boldsymbol{Q})\} \tag{13}$$

W is called the null space (Jaeckle and MacGregor, 1998) and is computed analytically by determining the kernel of the loading matrix $\boldsymbol{Q}$. All parameter combinations laying on the null space allow attaining $\boldsymbol{K}^*$ for the given **u**. Practically, the null space can be used to define the boundaries of the acceptable region, i.e., the possible combinations of parameters values leading exactly to the upper and lower limits for the KI of interest.

Recalling the illustrative example of Section 3.1.1, we can now relate the three model parameters to the correspondent response variable ($N_K = 1$). The projections of the calibration samples onto the score space (light blue circles) and the KS boundary (green ellipse) are shown in Fig. 5a for $B = 2$. Then, the PLS model is inverted (Fig. 5b) to determine the set of parameters combinations that guarantee that the prediction of the KI of interest is within the range of desired tolerance, for the given control variables. Since $B > N_K$, multiple solutions of the model inversion problem exist and are identified by the null space, which has dimension $B - N_K = 1$ (i.e., a straight line). Moreover, since the quality target is described by inequality constraints, three null spaces can be computed, i.e., one for the upper bound (upper black line), one for the target (central black line), and one for the lower bound (lower black line). Note that the null space corresponding to the target is not strictly necessary, because we are only interested in the acceptability region defined by the lower and upper bounds. We believe, however, that its representation is useful to the user in order to assess how close to the target the expected prediction is. Direct solutions $\boldsymbol{t}_{new}$ are also shown as black points, while the KI acceptability region corresponds to the grey area.

Estimated uncertainties of model parameters, which we express using the corresponding confidence intervals (CIs), can be projected onto the PLS model. Practically, if at least one projection falls outside the KI acceptability region (Fig. 5c), new experiments are needed to increase the parameter precision. No improvement is required if all uncertainties fall inside the KIs acceptability region (=stop criterion), as in Fig. 5d.

Although the methodology is based on a linear PLS model, i.e., an approximation of the rigorous model, the advantage is a graphical representation of both parameter uncertainty and KIs in a common latent space, allowing for an easier interpretation of the results.

3.3. Design and execution of experiments

Parameters are estimated on the basis of data obtained from experiments and, thus, they should be planned so as to facilitate the estimation activity. A set of experimental trials can be planned by simply acting on the experimental settings; the more common approaches are the full or fractional factorial DoE, where only a specific subset of control variables is varied (Box et al., 1978). These approaches are typically called "black box" experiment design methods because the analysis is carried out on the set of experimental conditions and measurements without considering any formal relationship between them. If a model is already available, MBDoE can represent an advantage since it exploits prior knowledge provided by the model (Franceschini and Macchietto, 2008). MBDoE techniques have been successfully applied in the pharmaceutical industrial environment, e.g., for model calibration in freeze-drying (De-Luca et al., 2020) or in the manufacturing of pharmaceutical agents (Shahmohammadi and McAuley, 2020). Since a model is available, MBDoE will represent the standard approach in this work. Details regarding the mathematical formulation of the MBDoE optimization problem are reported in Appendix A.1. Note that the design of experiments task can be tailored to gain the information relevant for the estimation of the most significant parameters only, i.e., those parameters that have the greatest impact on the prediction of KIs, as will be done in this study. However, other methods can also be implemented for model identification (Ljung, 1986).

The new experiment is, then, executed setting the experimental conditions planned by design.

3.4. Parameter estimation

The final step is the estimation of model parameters through available experimental measurements. Several methods can be adopted (e.g., least squares or a Bayesian estimator (Sorenson, 1980)). Here, a maximum likelihood approach is used (Bard, 1974). Since the method allows obtaining information on the conditioned probability distribution of the final estimate, we exploited it to compute the estimates of model parameters, and assess *a-posteriori* statistics. The mathematical formulation of the maximum likelihood function and the testing methods to assess *a-posteriori* statistics are reported in Appendix A.2.

4. Case study

We consider a DC systems model as illustrated in Fig. 1. The model is built under the following assumptions.

1. *Consistent/perfect blending.* API and excipients powders are perfectly mixed and have fixed quantities in the powder blend over the entire batch. This implies we can omit the feeding and blending unit operations and that blend, content uniformity and tablet weight variability can be omitted from this study.
2. *Dissolution test method.* The analytical method used to measure the in-vitro dissolution profile of the API is discriminatory, meaning that the method can capture changes in other input factors that could impact the dissolution performance (i.e., different input setpoints lead to different dissolution profiles). This implies that previous work has been carried out for the development of the analytical method for dissolution testing for this specific API, i.e., high performance liquid chromatography (HPLC) or ultraviolet (UV) spectroscopy calibration has been performed.

The model equations are presented in the following.

4.1. Model for the tablet press unit operation

The variation in tablet solid fraction caused by the compaction pressure, P [MPa], exerted by the press is expressed according to Kawakita and Lüdde (1971):

$$sf = \frac{a_{sf}\left(1 + b_{sf}P\right)}{1 + a_{sf}b_{sf}P} \tag{14}$$

where sf [–] is the attained tablet solid fraction, while a_{sf} [–] and b_{sf} [MPa^{-1}] are model parameters to be estimated. The Kushner (2012) equation is used to relate the effect of the extent of lubrication K [dm] attained in the upstream powder blending on the tablet tensile strength:

$$TS = TS_0((1-\beta) + \beta\exp(-\gamma K)) \tag{15}$$

where TS [MPa] is the tensile strength, TS_0 [MPa] is the tensile strength at zero porosity, γ[dm^{-1}] is the lubrication rate constant, and β [–] is the total fraction of tensile strength that can be lost due to lubrication. Given that the original Kushner equation valid only for solid fraction $sf = 0.85$, the empirical model by Nassar et al. (2021) is used to account for the dependence of the Kushner parameters of Eq. (15) on the attained tablet solid fraction:

$$TS_0 = a_1\exp(b_1(1-sf)), \tag{16}$$

$$\beta = a_2(1-sf) + b_2 \tag{17}$$

Seven model parameters associated with the tablet press need to be estimated:a_{sf} [–], b_{sf} [MPa^{-1}], γ[dm^{-1}], a_1 [MPa], b_1 [–], a_2[–], b_2 [–].

Note that in the industrial practice, it is more typical to have measurements of tablet hardness instead of tablet tensile strength. However, TS can be easily derived based on tablet geometry; for capsule-shaped tablets, the relation proposed by Pitt and Heasley (2013) can be used:

$$TS = \frac{2}{3}\left[\frac{10F_L}{\pi d^2\left(2.84\frac{T_t}{d} - 0.126\frac{T_t}{w} + 3.15\frac{w}{d} + 0.01\right)}\right] \tag{18}$$

where F is the tablet hardness [MN], T_t [m] is the tablet thickness, d[m] is the length of the short axis of the tablet, and w [m] is the wall height of the tablet.

4.2. Model for the tablet disintegration test

Both erosion and swelling mechanisms are considered (Markl et al., 2017).

Erosion is described as:

$$V_c = \left(H_{coat} - \dot{\epsilon}t\right)A_t, \tag{19}$$

where V_c [m^3] is the coating volume varying with time t [s], H_{coat} [m] is the thickness of the coating layer, A_t [m^2] is the tablet surface area, and $\dot{\epsilon}$ [m/s] is the constant erosion rate.

The dynamic evolution of the penetration depth due to swelling is modelled as:

$$\frac{dP_d}{dt} = \left(\frac{P}{F_L/A_t}\right)^{n\left(T_{t/2}-P_d\right)/T_{t/2}}\left[\frac{d_h^2\varepsilon}{S_p\tau_{or}^2\mu P_d}\right]p_c, \tag{20}$$

where P_d [m] is the water penetration depth, d_h [m] the tablet hydraulic diameter, τ_{or} [–] the average tablet tortuosity, μ [Pa s] the liquid viscosity, p_c [Pa] the capillary pressure. S_p [–] and n [–] are formulation-dependent model parameters to be estimated. Parameter n can be estimated if dynamic penetration depth data are available, e.g., via terahertz (THz) spectroscopy. However, this is a relatively new technology in pharmaceutical industrial environments compared to a standard end-point disintegration time test (USP ⟨7 0 1⟩ (2011)). If only end-point disintegration data are available, the term $\left(\frac{P}{F_L/A_t}\right)^{n\left(T_{t/2}-P_d\right)/T_{t/2}}$ in Eq. (20) can be replaced with a lumped parameter ω [–] which can be fitted. $T_{t/2}$ [m] is the time-dependent half tablet thickness, and ε [–] represents the average porosity of the swollen product. The stress due to tablet expansion from swelling is defined according to Peppas and Colombo (1989):

$$\tau = -TS + C_2w_l + C_3\sqrt{w_l}, \tag{21}$$

where τ [MPa] is the total stress, w_l [–] is the liquid content in the tablet, and C_2 [MPa] and C_3 [MPa] are model parameters to be estimated.

From τ, we can compute ε to be included in Eq. (21):

$$\tau = \frac{G_0\exp\left(-\frac{E\varepsilon}{1-\varepsilon}\right)\lambda t}{T_{t/2}}, \tag{22}$$

with G_0 [MPa] and E [–] elastic constants from literature, and λ [s^{-1}] the swelling rate (Kuentz and Leuenberger, 1998). The disintegration time is defined as the time for which the tablet stops disintegrating, i.e., $dP_d/dt = 0$.

Five model parameters associated with the disintegration test unit need to be estimated: C_2, C_3, $\dot{\epsilon}$, n, S_p.

4.3. Model for the in vitro dissolution test

We describe the rate of dissolution and the dissolution profile of each component of the formulation through a population balance approach (Wilson et al., 2012). For this purpose, the elements that constitute the formulation are divided in: (*i*) API, (*ii*) soluble excipients, (*iii*) insoluble excipients.

Focusing only on the API, the dynamic evolution of the number of particles, N_{API}, is modelled as:

$$\frac{\partial N_{API}}{\partial t} = B_{API}\delta\left(l - l_{0,API}\right) + R_{API,l}\frac{\partial N_{API}}{\partial l}, \tag{23}$$

where B_{API} is the rate of release of API from the tablet, l is the particle size at given t, $l_{0,API}$ is the particle size at the beginning of the process, and $R_{API,l}$ is the particle dissolution coefficient; δ is the Dirac delta function. Assuming particles have porosity much lower than the tablet porosity, and released particles are monodispersed with a fixed average diameter $l_{0,API}$, B_{API} can be expressed as:

$$B_{API} = \frac{1}{\rho_p}\left(\frac{x_{API}}{\phi l_{0,API}^3}\right)\frac{dM_t}{dt}, \tag{24}$$

where x_{API} is the mass fraction of API,ρ_p is the density of particles, and ϕ is the shape factor of particles ($\phi = \pi/6$ for spherical particles); dM_t/dt represents the dynamic evolution of the tablet mass M_t.

$R_{API,l}$ can be described by:

$$R_{API,l} = k_{API}\left(c_{sat} - c_{API}\right)^{n_{API}}, \tag{25}$$

where k_{API} is the mass transfer coefficient of API, n_{API} is the order of dissolution, and c_{sat} and c_{API} are the API saturation concentration and the bulk concentration, respectively. The extent of dissolution is typically described as the percentage of label content %LC:

$$\%LC = 100\frac{c_{API}V_m}{x_{API}M_{t,0}}, \tag{26}$$

where V_m is the liquid volume in the test vessel, and $M_{t,0}$ is the initial mass of the tablet.

Two model parameters need to be estimated for the in vitro tablet dissolution test unit: k_{API}, and n_{API}.

Table 1
Nominal and initial guess values of model parameters.

Parameter	Units	Nominal	Initial guess
Tablet press unit operation			
a_1	MPa	11.04	12.01
a_2	–	1.091	1.363
a_{sf}	–	0.463	0.533
b_1	–	−8.202	−9.005
b_2	–	0.326	0.419
b_{sf}	MPa^{-1}	2.460×10^{-2}	1.990×10^{-2}
γ	dm^{-1}	1.211×10^{-3}	1.685×10^{-3}
Tablet disintegration test unit			
C_2	MPa	1.000×10^{2}	1.500×10^{2}
C_3	MPa	1.000×10^{2}	1.500×10^{2}
$\dot{\epsilon}$	m/s	1.000×10^{-3}	8.125×10^{-4}
n	–	0.900	0.978
S_P	–	0.524	0.598
In vitro *dissolution test unit*			
k_{API}	$(m^3\ kg^{-1})^n API\ s^{-1}$	2.300×10^{-12}	2.615×10^{-12}
n_{API}	–	1.00	1.00

Table 2
Tablet press unit operation. Diagnostics of the PCA model: eigenvalues and explained variance per principal component.

PC	Eigenvalues	Explained Variance (%)	Cumulative Variance (%)
1	1.845	26.37	26.37
2	1.062	15.17	41.54
3	1.025	14.65	56.19
4	1.009	14.42	70.61
5	0.993	14.18	84.79
6	0.990	14.14	98.93
7	7.494×10^{-2}	1.07	100

4.4. Product quality and performance assessment

As shown in Fig. 1, three KIs are here considered (which correspond to each sub-model output): tensile strength, target disintegration time, and API dissolution profile. Their acceptability limits are discussed in the following.

We consider an immediate release (IR) tablet, with a target *TS* of 2 MPa. According to acceptable error ranges suggested by Nassar et al. (2021), we set ± 0.2 MPa as the admissible tolerance with respect to the *TS* target value.

The target disintegration time is assumed to be 4 min. According to the USP ⟨7 0 1⟩ (2011) disintegration test specifications, the time limit for the formulation to completely disintegrate is 5 min; therefore, we set ± 1 min as the admissible tolerance with respect to the target value of the disintegration time.

The dissolution profile is monitored through the prediction of %*LC*. %$LC = 80\%$ at $t = 25$ min is a possible specification value for an immediate release tablet; however, the actual specification will depend on the specific product. We set − 15 %*LC* as the admissible tolerance with respect to the target value; no overestimation is accepted for a conservative analysis.

Table 3
Tablet press unit operation. Comparison of different PCA models with respect to the reference case i ($P_i = 241$ MPa, $K_i = 1090$ dm) through the similarity factor $S^{\lambda}_{i,j}$.

Operating conditions j		
P_j[MPa]	K_j[dm]	$S^{\lambda}_{i,j}$
150	800	0.990
250	2000	0.986
350	900	0.976
400	1500	0.974

Table 4
Tablet disintegration test unit. Diagnostics of the PCA model: eigenvalues and explained variance per principal component.

PC	Eigenvalues	Explained Variance (%)	Cumulative Variance (%)
1	2.227	18.56	18.56
2	1.306	10.88	29.44
3	1.081	9.01	38.45
4	1.043	8.69	47.14
5	1.029	8.58	55.72
6	1.009	8.41	64.13
7	0.999	8.33	72.46
8	0.985	8.21	80.67
9	0.964	8.03	88.70
10	0.923	7.70	96.40
11	0.372	3.09	99.49
12	6.149×10^{-2}	0.51	100

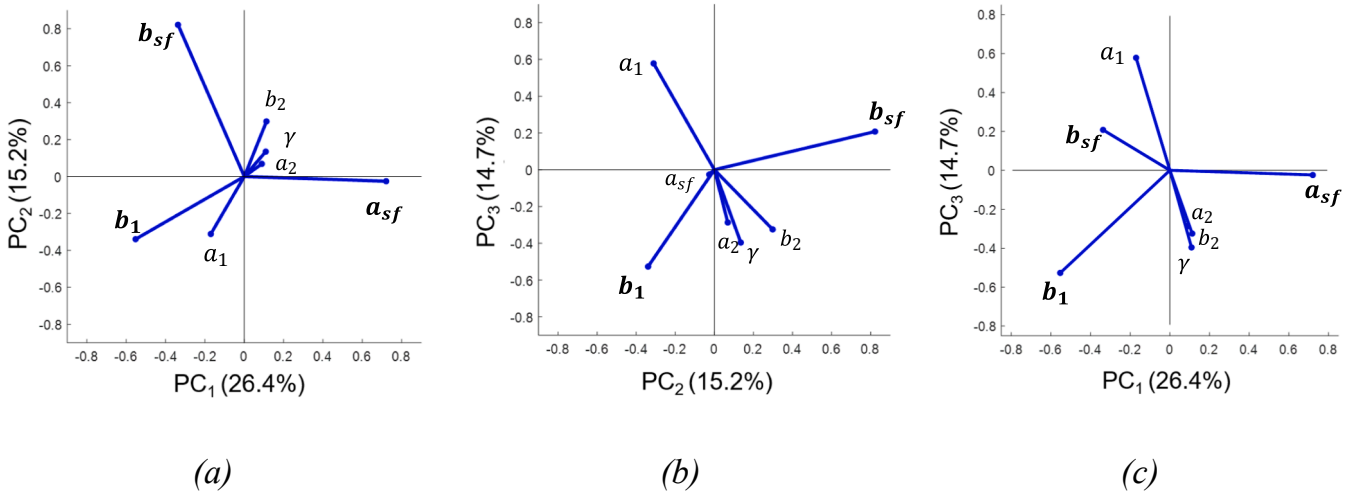

Fig. 6. Results of PCA applied to the Tablet press unit operation: plot of loadings for (a) PC1 vs. PC2, (b) PC2 vs. PC3 and (c) PC1 vs. PC3. The most influential parameters are in boldface.

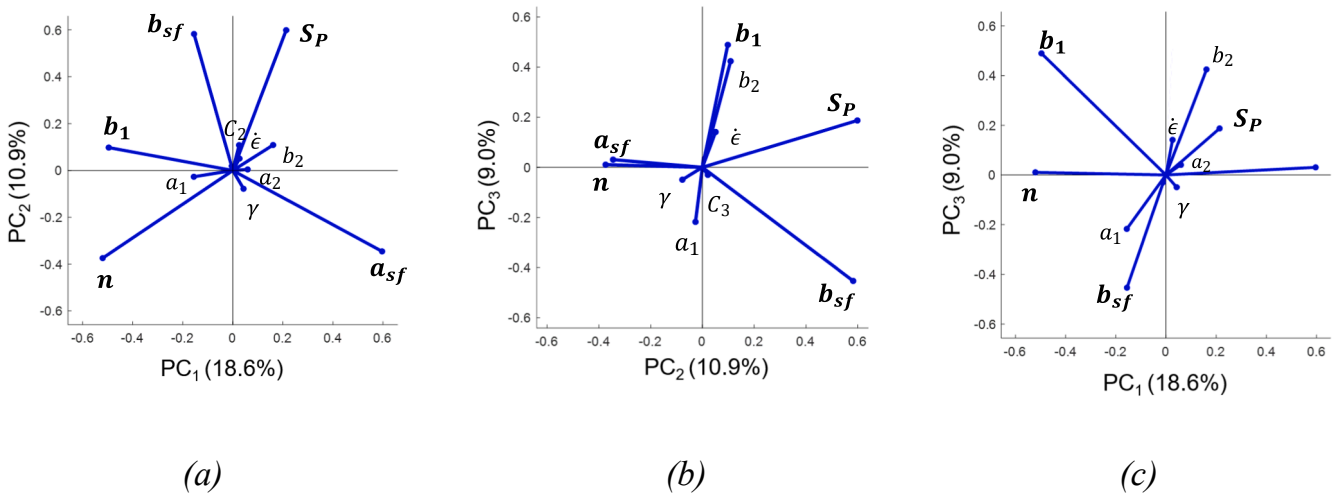

Fig. 7. Results of PCA applied to the tablet disintegration test unit: plot of loadings for (a) PC1 vs. PC2, (b) PC2 vs. PC3, and (c) PC1 vs. PC3. The most influential parameters are in boldface.

Table 5
Tablet disintegration test unit. Comparison of different PCA models with respect to the reference case i ($P_i = 223$ MPa, $K_i = 990$ dm) through the similarity factor $S^{\lambda}_{i,j}$.

Operating conditions j		
P[MPa]	K[dm]	$S^{\lambda}_{i,j}$
150	800	0.998
250	2000	0.999
350	900	0.997
400	1500	0.998

Table 6
In vitro dissolution test unit. Diagnostics of the PCA model: eigenvalues and explained variance per principal component.

PC	**Eigenvalues**	**Explained Variance (%)**	**Cumulative Variance (%)**
1	1.784	12.74	12.74
2	1.235	8.82	21.56
3	1.174	8.39	29.95
4	1.116	7.97	37.92
5	1.095	7.82	45.74
6	1.064	7.60	53.34
7	1.036	7.40	60.74
8	0.972	6.94	67.68
9	0.963	6.88	74.56
10	0.869	6.21	80.77
11	0.814	5.81	86.58
12	0.771	5.51	92.09
13	0.751	5.37	97.46
14	0.355	2.54	100

5. Results

The framework is assessed by means of an in-silico case study. The *process* is represented by the systems model with parameters at nominal values (Table 1) as retrieved from the literature (Peppas and Colombo, 1989; Nassar et al., 2021). The *model* is represented by the systems model with initial guesses for the parameters as in Table 1. We assume that the initial parameter uncertainty is equal to ± 50% of their guess values (uniform distribution).

5.1. Model identifiability and parameters ranking

5.1.1. Tablet press unit operation

We first focus on the tablet press unit operation, where the KI of interest is the tensile strength *TS*. Assuming that the initial parameter uncertainty is equal to ± 50% of their initial values (Table 1), we generate a set of 1×10^4 parameters combinations using the MC method. Control variables (i.e., the tablet press model inputs P and K, which are time-invariant variables) are fixed so that the predicted *TS* based on current parameter values is equal to the target value (i.e., $P = 241$ MPa and $K = 1090$ dm).

We set $A = 4$ (Table 2); however, note that also PC$_5$ and PC$_6$ capture a significant amount of information, i.e., explained variance. The model parameters are plotted as points using their loadings on the PC coordinates; results are shown for PC$_1$ vs. PC$_2$ (Fig. 6a), PC$_2$ vs. PC$_3$ (Fig. 6b), and PC$_1$ vs. PC$_3$ (Fig. 6c). The case PC$_3$ vs. PC$_4$ is reported in the Supplementary Material (Section S.3.1). Parameter a_{sf} can be recognized as the most influential one by simply observing the length of its projected loading along PC$_1$ (Fig. 6a), which is the component capturing the highest percentage of variance (Table 2). Following the same rationale, parameters b_1 and b_{sf} are very influential too, whereas a_2 and γ are barely affecting *TS*. It can be also noted that a_2, b_2, and γ are correlated, and are anticorrelated to a_1, (i.e., inversely correlated). Difficulties in their estimation are expected – particularly for the less influential parameters (e.g., a_2 may not attain a statistically satisfactory precision).

The influence of the operating variables on the correlation structure of model parameters is assessed using the similarity factor: the current PCA model (whose control variables are referred to as reference operating conditions i) is compared to other four PCA models (whose operating conditions j are reported in Table 3). In all cases, $S^{\lambda}_{i,j}$ values are very close to 1, suggesting that the correlation structure and the ranking of model parameters is scarcely dependent on the operating conditions. Therefore, in this case study, it is unnecessary to perform the PCA procedure again when operating conditions change along subsequent iterations of the procedure in Fig. 3, as they barely affect the interaction between parameters and their relative ranking.

5.1.2. Tablet disintegration test unit

The prediction fidelity of the disintegration time not only depends on the parameters of the model for the tablet disintegration test unit, but also on the parameters of the model for the tablet press unit operation; therefore, all the parameters of the models of the two units should be

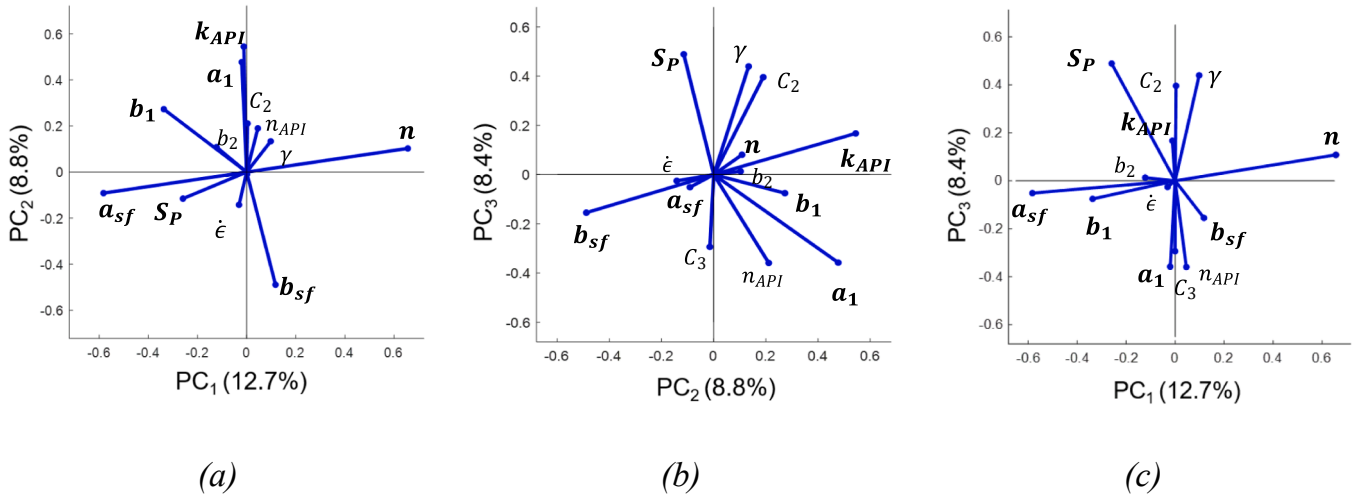

Fig. 8. Results of PCA applied to the In vitro dissolution test unit: plot of loadings for (a) PC1 vs. PC2, (b) PC2 vs. PC3, and (c) PC1 vs. PC3. The most influential parameters are in boldface.

Table 7
In vitro dissolution test unit. Comparison of different PCA models with respect to the reference case i ($P_i = 223$ MPa, $K_i = 990$ dm) through the similarity factor $S^{\lambda}_{i,j}$.

Operating conditions j		
P[MPa]	K[dm]	$S^{\lambda}_{i,j}$
150	800	0.983
250	2000	0.987
350	900	0.986
400	1500	0.981

Table 8
Tablet press unit operation. Diagnostics of the PLS model: explained variance per latent variable.

LV	Explained Variance (%)	Cumulative Variance (%)
1	66.76	66.76
2	29.11	95.86
3	0.33	96.19

considered when the PCA procedure is performed. Given the initial parameter guesses (Table 1) and assuming that their uncertainty is equal to ± 50% of their initial values, we generate a set of 1×10^4 parameters combinations and we find the control variables (i.e., the tablet press model inputs P and K) so that the predicted disintegration time is equal to 4 min (i.e., $P = 223$ MPa and $K = 990$ dm). According to the eigenvalue greater than one rule, at least $A = 6$ PCs are required to capture enough variance of data (Table 4). For the sake of clarity, plots of loadings are shown for PC_1 vs. PC_2 (Fig. 7a), PC_2 vs. PC_3 (Fig. 7b), and PC_1 vs. PC_3 (Fig. 7c). Plots of loadings for PC_3 vs. PC_4, PC_4 vs. PC_5, and PC_5 vs. PC_6 are reported in the Supplementary Material (Section S.3.2). Several influential parameters can be noticed: n, a_{sf}, b_1, b_{sf}, and S_P. The variance of n, a_{sf}, and b_1 is mainly captured by PC_1, while the variance of b_{sf} and S_P is better explained by PC_2. Note that a_{sf}, b_1 and b_{sf} belong to the tablet press model. As far as the correlation of model parameters is concerned, the parameters related to the tablet press unit operation behave similarly to the case in Section 5.1.1; parameters n and S_P appear to be anticorrelated.

As done for the model for the tablet press unit operation, the influence of the operating variables on the correlation structure of model parameters is assessed using the similarity factor (different operating conditions are reported in Table 5). The resulting $S^{\lambda}_{i,j}$ values are close to 1, implying that it is unnecessary to perform the PCA procedure again when operating conditions change.

5.1.3. *In vitro dissolution test unit*

The prediction fidelity of the API dissolution profile depends not only on the parameters of the model for the in vitro dissolution test unit, but also on the parameters of the models for the tablet press unit operation and the tablet disintegration test unit; therefore, all parameters of the three-unit models should be considered when the PCA procedure is performed. Given the initial parameter guesses (Table 1) and assuming that their uncertainty is equal to ± 50% of their initial values, we generate a set of 1×10^4 parameters combinations and we find the control variables (i.e., the tablet press model inputs P and K) so that the predicted dissolution of API attained in 25 min is equal to 80 %LC (i.e., $P = 223$ MPa and$K = 990$ dm). In this case, at least $A = 7$ PCs are required to capture enough variance in the data (Table 6). Plots of loadings are shown for PC_1 vs. PC_2 (Fig. 8a), PC_2 vs. PC_3 (Fig. 8b), and PC_1 vs. PC_3 (Fig. 8c). Plots of loadings for PC_3 vs. PC_4, PC_4 vs. PC_5, PC_5 vs. PC_6, and PC_6 vs. PC_7 are reported in the Supplementary Material (Section S.3.3). It can be observed that parameters with the greatest loadings are a_{sf}, b_{sf}, a_1,b_1 (tablet press model), n, S_P (tablet disintegration model), and k_{API} (in vitro dissolution model). The correlation structure among the parameters of previous units does not change; k_{API} is anticorrelated with parameter b_{sf}.

We finally assess the influence of the operating conditions (Table 7) on the correlation structure of model parameters. In this case, too, resulting $S^{\lambda}_{i,j}$ values are very close to 1, and therefore, we do not need to repeat the PCA procedure for different operating conditions.

5.2. *Assessment of parameter uncertainty on the fidelity of model predictions*

Both the modular and global approaches were implemented and discussed in the following sections.

5.2.1. *Modular approach*

All the KIs are targeted sequentially, while model parameters are estimated on a unit operation basis.

We first focus on the model for the tablet press unit operation, where the KI of interest is TS ($N_K = 1$). The relationship between model parameters and the predicted TS is assessed by building a PLS model, according to Section 3.2. We verified that two LVs ($B = 2$) are sufficient to

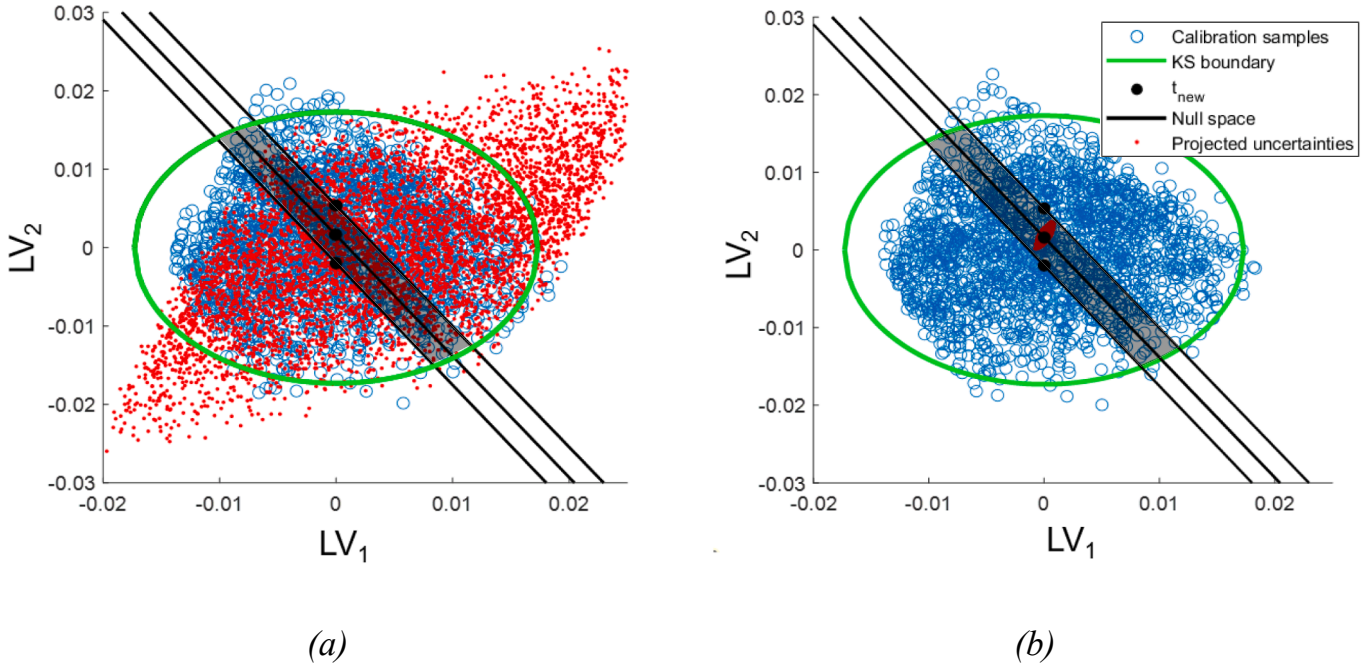

Fig. 9. Modular approach. Application of the proposed methodology to the tablet press unit operation for the assessment of the tensile strength: (a) initial iteration, (b) final iteration where all projected uncertainties (red points) fall inside the KI acceptability region (grey area).

Table 9
Tablet press unit operation. Estimated values of model parameters with their 95% CIs and t-values. * = precision is not statistically satisfactory.

Parameter	Units	Nominal	Estimated	95% CI	*t*-value
a_1	MPa	11.04	12.09	0.516	23.45
a_2	–	1.091	0.545	0.406	1.35*
a_{sf}	–	0.463	0.427	5.296×10^{-3}	80.63
b_1	–	−8.202	−8.363	0.221	37.85
b_2	–	0.326	0.121	5.561×10^{-2}	2.18
b_{sf}	MPa^{-1}	2.460×10^{-2}	2.010×10^{-2}	2.209×10^{-4}	1.255×10^{2}
γ	dm^{-1}	1.211×10^{-3}	1.140×10^{-3}	5.527×10^{-4}	2.75
					$t_{ref} = 1.690$

Table 10
Tablet disintegration test unit. Diagnostics of the PLS model: explained variance per latent variable.

LV	Explained Variance (%)	Cumulative Variance (%)
1	59.05	59.05
2	9.03	68.08
3	0.41	68.49

capture enough variance of the data in $\mathscr{K}$ (Table 8). As $B > N_K$, multiple solutions of the model inversion problem exist and are identified by the null space, that has dimension $B - N_K = 1$, i.e., a straight line. More precisely, since the quality target is described by inequality constraints, three null spaces can be computed to define the KI acceptability region, i.e., one for the upper bound, one for the target, and one for the lower bound. By assessing the effect of the initial model parameter uncertainty on the KI, it can be observed that the prediction fidelity requirement for the *TS* cannot be satisfied, as many projections lie outside of the KI acceptability region (Fig. 9a).

MBDoE is applied to increase the precision of the three most relevant parameters only, namely a_{sf}, b_1, and b_{sf}, by acting on design variable P and K. Data of *TS* are used to estimate all parameters for the current unit. Eight iterations are needed to reach the required model fidelity with respect to the *TS* prediction, i.e., eight experiments need to be performed for the tablet press unit operation. The available tablet press equipment may allow the operator to set more compression levels. This occurs for example when a compaction simulator is used that can collate multiple data for solid fraction for a tablet compressed from a single powder blend with lubrication extent K. In this case, the MBDoE procedure proposed by Cenci et al. (2022) can be exploited.

Results after the final iteration are shown in Fig. 9b: all projected uncertainties of parameter estimates fall inside the KI acceptability region. Table 9 reports the estimated values of model parameters with their 95% CIs and *t*-values. It can be observed that parameter a_2 does not require precise estimation – the corresponding *t*-value is smaller than the reference value and is associated to a large CI. Namely, there is no need of precisely estimating parameter a_2 to meet the *TS* specification, and therefore additional experimental effort can be saved. The result agrees with the outcome from the PCA analysis, where parameter a_2 was ranked as having little influence.

The next unit is the tablet disintegration test. It is assumed that the parametric precision attained for the tablet press model does not need any further improvement. A new PLS model is built with $B = 2$ LVs (Table 10). As a consequence, multiple solutions of the model inversion problem exist and are identified by the null space, that has dimension $B - N_K = 1$. Since the quality target is described by inequality constraints, three null spaces can be computed to define the acceptability region for the disintegration time, i.e., one for the upper bound, one for the target, and one for the lower bound. Given the initial model parameter uncertainty, we can observe that the prediction fidelity

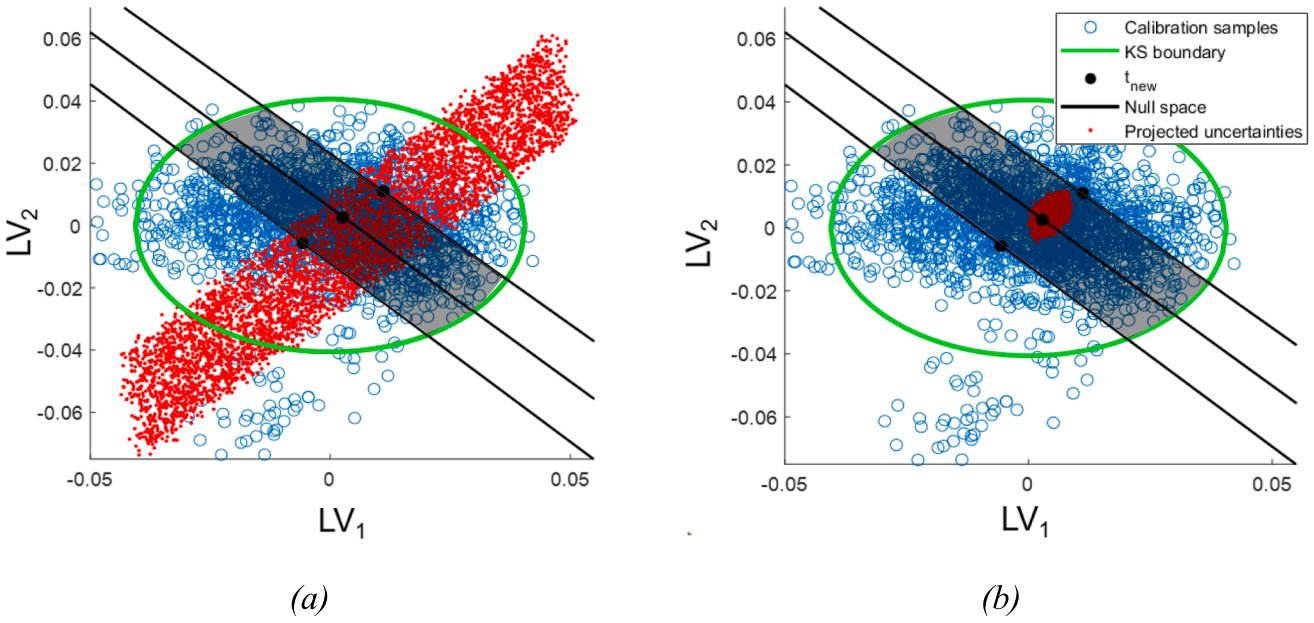

Fig. 10. Modular approach. Application of the proposed methodology to the tablet disintegration test unit for the assessment of the disintegration time: (a) initial iteration, (b) final iteration where all projected uncertainties (red points) fall inside the KI acceptability region (grey area).

Table 11
Tablet disintegration test unit. Estimated values of model parameters with their 95% CIs and t-values. † = 95% CI larger than ± 50% the parameter nominal value. * = precision is not statistically satisfactory.

Parameter	Units	Nominal	Estimated	95% CI	*t*-value
C_2	MPa	1.000×10^2	1.487×10^2	2.968×10^4 †	5.009×10^{-3} *
C_3	MPa	1.000×10^2	80.48	4.195×10^3 †	1.919×10^{-2} *
$\dot{\epsilon}$	m/s	1.000×10^{-3}	8.599×10^{-4}	0.147 †	5.291×10^{-3} *
n	–	0.900	0.935	2.738×10^{-2}	34.16
S_P	–	0.524	0.675	0.108	6.25 $t_{ref} = 1.647$

Table 12
In vitro dissolution test unit. Diagnostics of the PLS model: explained variance per latent variable.

LV	Explained Variance (%)	Cumulative Variance (%)
1	65.66	65.66
2	10.42	76.08
3	1.37	77.45

requirement for the disintegration time is not satisfied, as many projections lie outside of the KI acceptability region (Fig. 10a).

MBDoE is applied to increase the precision of the two relevant parameters n and S_P (see Section 5.1.2), by acting on the design variables K and P. Disintegration data are used to estimate all the parameters of the model for current unit. Five iterations (i.e., five experiments for the current unit) are necessary to reach the required model fidelity with respect to the KI prediction. Results after the final iteration are collected in Table 11. It can be observed that we need to precisely estimate only the influential model parameters S_P and n, whereas satisfactory statistical precision is not required for parameters C_2, C_3 and $\dot{\epsilon}$ – the corresponding t-values are smaller than the reference value, and the corresponding CIs exceed ± 50% of the parameter estimates. Fig. 10b shows that after the final iteration all projected uncertainties fall inside the KI acceptability region.

We finally move to the in vitro dissolution test unit. It is assumed that parameters concerning previous units need no improvement. A new PLS model is built with $B = 2$ LVs, as it is found that after the second LV there is no improvement in the total amount of explained variance of data in $\mathscr{X}$ (Table 12). The null space has dimension $B - N_K = 1$. Recalling that we set − 15 %LC as the admissible tolerance with respect to the target value, with no overestimation accepted, two null spaces can be computed to define the boundaries of the acceptability region for the %LC attained after 25 min, i.e., one for the lower bound, and one for the target. Given the initial model parameter uncertainty, the prediction fidelity requirement for the %LC at $t = 25$ min is not satisfied, as many projections lie outside of the KI acceptability region (Fig. 11a).

MBDoE is applied to increase the precision of the most influential parameter k_{API} (see Section 5.1.3), by acting on design variables P and K, and exploiting API dissolution data as measured output variable to estimate both k_{API} and n_{API}. One single iteration (i.e., one experiment for the current unit) is necessary to reach the required model fidelity with respect to the KI prediction, which is found to depend only on the value of k_{API}. Results in Table 13 show that we need to precisely estimate parameter k_{API}, while n_{API} is not precisely estimated from experimental measurements. This confirms that model identifiability may be unnecessary for the purpose of achieving model fidelity, as large uncertainty on parameter n_{API} does not lead to large uncertainty on the prediction of the API dissolution profile. Its value can be fixed based on similar systems in the literature (e.g., Bano et al., 2022). Fig. 11b shows that all projected uncertainties fall inside the KI acceptability region.

5.2.2. *Global approach focusing on all KIs simultaneously*

In the global approach, all KIs are targeted simultaneously, and the parameters of all unit operation models are considered at the same time. The relationship between model parameters and all KIs predictions ($N_K = 3$) is assessed by building a PLS model, according to Section 3.2.

Four LVs capture a reasonable amount of variance of the data in $\mathscr{X}$ – from LV_5 onwards there is no significant improvement (Table 14), i.e., $B = 4$ can be suitable to represent the variability of predicted KIs (i.e., TS, disintegration time, and %LC at $t = 25$ min). Since the final product quality is defined by the independent control of each KI of interest, no

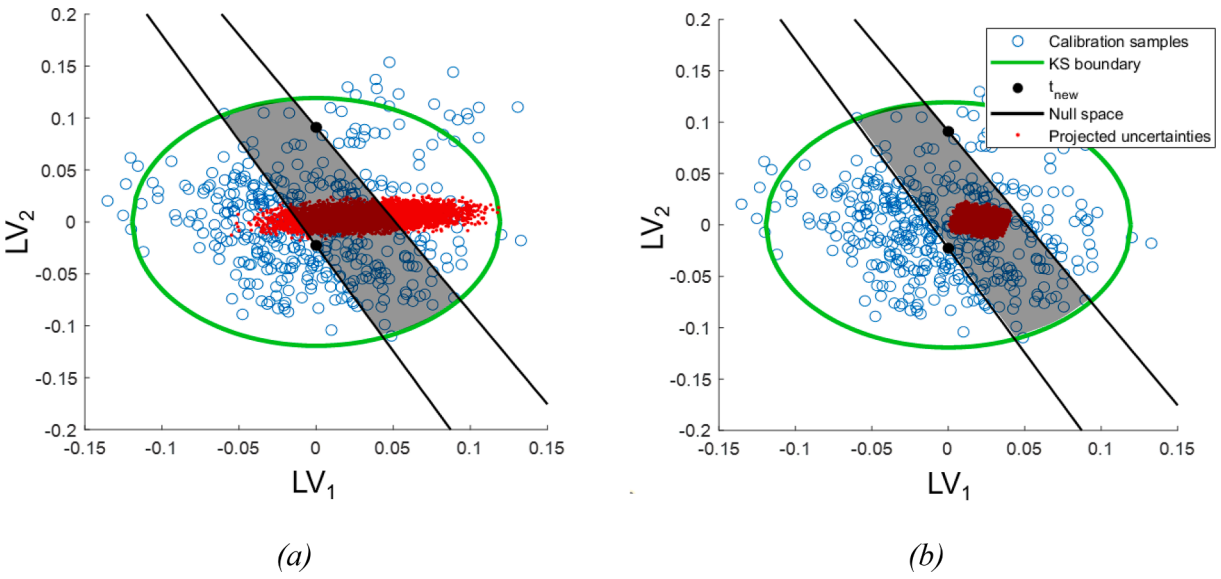

Fig. 11. Modular approach. Application of the proposed methodology to the in vitro dissolution test unit for the assessment of %*LC* attained in 25 min: (a) initial iteration, (b) final iteration where all projected uncertainties (red points) fall inside the KI acceptability region (grey area).

Table 13
In vitro dissolution test unit. Estimated values of model parameters with their 95% CIs and t-values. †† = parameter estimate equal to the upper bound (UB) of its estimation range.

Parameter	Units	Nominal	Estimated	95% CI	*t*-value
k_{API}	$(m^3\ kg^{-1})^n API\ s^{-1}$	2.300×10^{-12}	1.950×10^{-12}	6.189×10^{-17}	3.150×10^4
n_{API}	–	1.00	UB †† (=1.50)	–	–
					t_{ref} = 1.646

Table 14
Comprehensive approach focusing on all KIs simultaneously. Diagnostics of the PLS model: explained variance per latent variable.

LV	Explained Variance (%)	Cumulative Variance (%)
1	19.79	19.79
2	14.32	34.11
3	14.33	48.44
4	8.62	57.06
5	1.93	58.99

correlation among the outputs is considered in the PLS model inversion activity. Multiple solutions of the model inversion problem exist and are identified by the null space, that has dimension $B - N_K = 1$. Projected uncertainties of model parameters are shown on LV_1 vs. LV_2 (Fig. 12). Null spaces are computed considering the extreme values of the tolerance range for each KI; i.e., two null spaces are identified: one for the lower bounds and one for the upper bounds. Given the initial model parameter uncertainty, we can observe that the prediction fidelity requirement for the three model KIs is not met, as many projections lie outside the acceptability region (Fig. 12a). Therefore, MBDoE is applied to increase the precision of the most influential parameters towards the prediction of all model KIs, i.e.,a_1, a_{sf}, b_1, b_{sf} (tablet press model), n, S_P (tablet disintegration model), and k_{API} (in vitro dissolution model). Design variables K and P are again used in the MBDoE problem. For each iteration, three different experiments for the three model units need to be performed simultaneously. Measured output variables are *TS*, disintegration data, and API dissolution data.

Eight iterations are needed to reach the required fidelity of the KIs, i.e., eight experimental iterations would need to be performed for each unit simultaneously, i.e., 24 experiments altogether. Results after the final iteration are shown in Fig. 12b, where all projected uncertainties of parameter estimates fall inside the KI acceptability region. Table 15 reports the estimated values of model parameters with their 95% CIs and *t*-values. Observations regarding values of estimates and their precision are similar to the case discussed in Section 5.2.1. Parameters a_2 and n_{API} cannot be estimated from experimental measurements – their value is set to the lower and upper bounds of their estimation ranges, respectively. However, since they have little influence towards the considered KI predictions, there is no need to identify them. Parameters C_2, C_3 and $\dot{\epsilon}$, too, do not require precise estimation. All KI specifications are met, and no further experimental effort is required.

An additional case was analysed, where only one final KI, related to the in vitro dissolution unit (i.e., %*LC* at $t = 25$ min), was considered. Seven iterations are needed to reach the required model fidelity with respect to the %*LC*, i.e., seven experiments would need to be performed for each unit simultaneously, with a total of 21 experiments. The precision of model parameters is similar to results of the previous case studies. Details are reported in Section S.4 of the Supplementary material.

5.2.3. Computational details

All activities were performed on an Intel Core I7- 11850H CPU@2.50 GHz processor with 64.0 GB RAM. We used MATLAB® R2021b to perform MC simulations, and to construct PCA and PLS models. We used gPROMS v. 7.0.7 for process simulation, to implement MBDoE, and for parameter estimation.

Performance of MC simulations is the most time-demanding step of the entire procedure, and depends on the model complexity and on the number of parameters to be considered. The required time to perform 10^4 MC simulations was: (i) few seconds for the model for the tablet press unit operation, (ii) ~ 30 min for the model for the tablet disintegration test unit, (iii) ~ 90 min for the model for the in vitro dissolution test unit. Therefore, even if the modular approach consists of more iterations than the global one (14 vs. 8), its computational time is considerably shorter (~4h vs. ~ 12 h).

All other activities require little computational time. PCA and PLS derive from MC results, and only some seconds are needed for their construction. Implementation of MBDoE and parameter estimation both

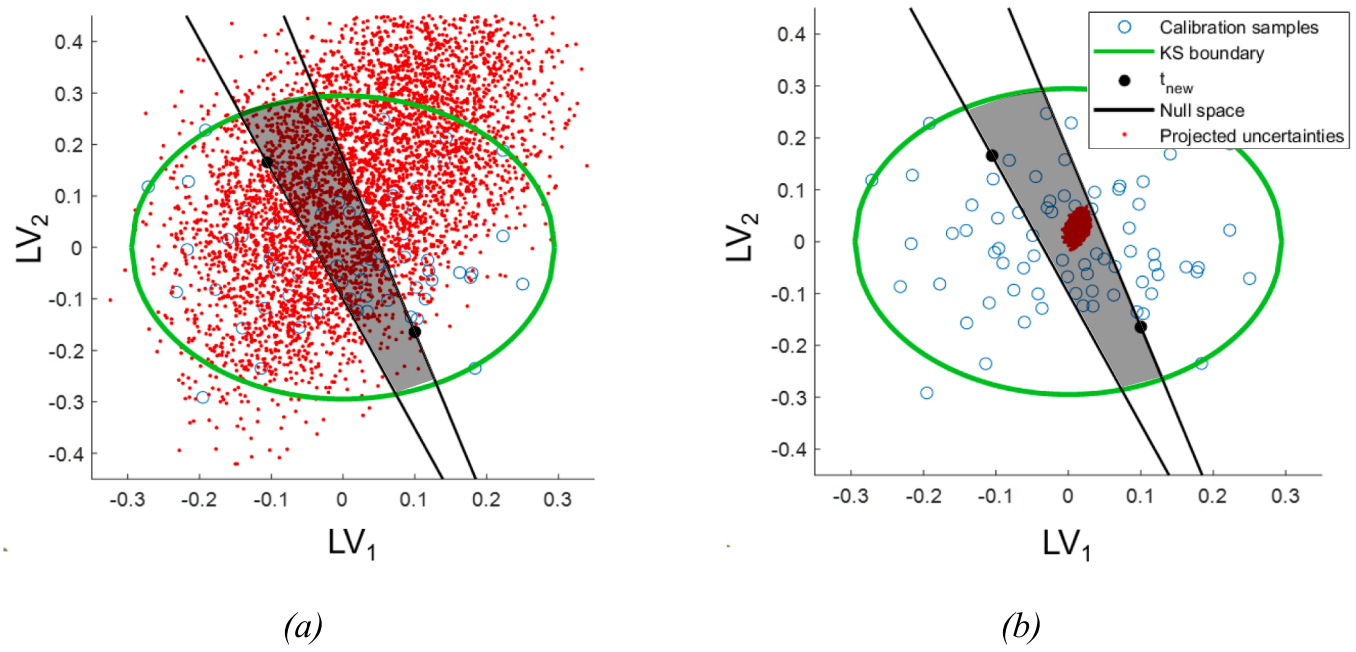

Fig. 12. Global approach. Application of the proposed methodology to the DC systems model for the assessment of all model KIs: (a) initial iteration, (b) final iteration where all projected uncertainties (red points) fall inside the KI acceptability region (grey area).

Table 15
Global approach focusing on all KIs simultaneously. Estimated values of model parameters with their 95% CIs and t-values. † = 95% CI larger than ± 50% the parameter nominal value. ††= parameter estimate equal to the lower bound (LB) or upper bound (UB) of its estimation range. * = precision is not statistically satisfactory.

Parameter	Units	Nominal	Estimated	95% CI	*t*-value
a_1	MPa	11.04	10.04	0.499	20.12
a_2	–	1.091	LB†† (=0.500)	–	–
a_{sf}	–	0.463	0.458	2.029×10^{-3}	2.259×10^{2}
b_1	–	−8.202	−8.238	6.062×10^{-2}	1.359×10^{2}
b_2	–	0.326	0.320	7.991×10^{-3}	40.09
b_{sf}	MPa^{-1}	2.460×10^{-2}	2.537×10^{-2}	2.491×10^{-4}	1.018×10^{2}
γ	dm^{-1}	1.211×10^{-3}	1.106×10^{-3}	4.934×10^{-4}	22.42
					$t_{ref} = 1.690$
C_2	MPa	1.000×10^{2}	1.499×10^{2}	6.553×10^{4} †	2.289×10^{-3} *
C_3	MPa	1.000×10^{2}	1.497×10^{2}	1.471×10^{2} †	0.102 *
$\dot{\epsilon}$	m/s	1.000×10^{-3}	9.973×10^{-3}	3.084×10^{-2} †	2.271×10^{-2} *
n	–	0.900	0.901	3.443×10^{-2}	26.17
S_P	–	0.524	0.491	9.818×10^{-2}	5.00
					$t_{ref} = 1.647$
k_{API}	$(m^3\ kg^{-1})^n API\ s^{-1}$	2.300×10^{-12}	1.982×10^{-12}	3.797×10^{-17}	5.217×10^{4}
n_{API}	–	1.00	UB †† (=1.50)	–	–
					$t_{ref} = 1.646$

require short time, in between seconds and few minutes (1–2 min) for all sub-models. Process simulation demands only few seconds for all submodels.

5.3. Discussion

The case study demonstrated the effectiveness of the proposed approach for systematic evaluation of pharmaceutical process systems models. Here are some additional comments:

- In this case study, the modular approach appears to be more efficient than the global one as the overall number of experimental runs is less (14 vs. 24). However, a greater number of iterations – i.e., sequences of design of experiments, experimental runs and parameter estimation tasks – is needed (14 vs. 8), although with a lower computational burden. However, note that in the modular approach, in case a KI cannot be predicted with the required fidelity, it is more difficult to assess whether more experiments are needed on the unit being investigated (i.e., the unit outputting the KI) or, conversely, more experiments should be carried out in one or more of the previous units in order to achieve better estimates of the parameters in those units; uninformative experiments may be carried out before one realizes that higher precision is required elsewhere;
- The management of the experimental campaign is generally easier in the modular approach since experiments can be organized unit by unit, thus facilitating a more efficient scheduling;
- Although not shown here, we verified that a tighter tolerance on the KIs can increase the required experimental effort dramatically; in such a case, reducing the experimental cost can be the priority requirement, and optimizing the experimental effort in a modular approach may be a difficult task. Also, note that if a larger parameter uncertainty is present (e.g., more than 50% of the nominal value),

the required number of iterations and the computational burden may considerably increase. The experimental effort will increase accordingly;
- Graphical approaches were developed to assess the importance of different parameters in the prediction of KIs, and to quantify the impact of estimation uncertainty on model fidelity. We believe they are particularly effective in making the outcomes of some required theoretical steps clear and comprehensible to everyone. On the other hand, we recognize that in complex models with large number of parameters, their implementation may become cumbersome and computationally demanding; other techniques may be adopted within the proposed framework;
- Our methodology relies on the assumption that the only mismatch between model and process depends on the parameter values (parametric mismatch). In a real-life situation, a structural mismatch between the model and the process often exists. However, the presence of relevant discrepancies between model and process would require changes on model equations or the introduction of correction approaches (e.g., via hybrid modeling). How structural mismatch can be handled effectively is a subject of further investigation.
- Both PCA and PLS models are built by considering the relation between model parameters and outputs, while there is no linearization of the input–output map. However, if the model is strongly nonlinear with respect to the relationship between parameters and KIs, the uncertainty region may be represented ineffectively.

6. Conclusions

A systematic procedure for the assessment of the prediction fidelity of quantitative (systems) models has been proposed and efficiently implemented. Our study paves the way to the systematic use of standardized approaches for model evaluation, and aims at enhancing model-based development for pharmaceutical manufacturing processes. Referring to the three questions considered in the introduction, it can be stated:

(i) A methodology has been proposed to evaluate uncertainty in model predictions (model fidelity) by quantifying the impact of parameter precision; it can be exploited to ensure pre-set requirements on parameters in the prediction of model KIs;
(ii) It has been verified that it is generally unnecessary to estimate all model parameters in a statistically satisfactory way to satisfy the KI requirements; focusing only on the parameters that are ranked as (highly) influential is typically sufficient;
(iii) No clear rule could be postulated on whether a calibration on a unit basis or a global calibration approach is the most effective strategy for systems models; pros and cons should be evaluated on a case by case basis.

Future work will aim at testing the procedure experimentally, and at investigating model-process structural mismatch thoroughly.

CRediT authorship contribution statement

Margherita Geremia: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Samir Diab:** Conceptualization, Methodology, Writing – review & editing. **Charalampos Christodoulou:** Conceptualization, Software. **Gabriele Bano:** Conceptualization, Writing – review & editing. **Massimiliano Barolo:** Conceptualization, Methodology, Writing – review & editing. **Fabrizio Bezzo:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Fabrizio Bezzo reports financial support was provided by GlaxoSmithKline (GSK). Gabriele Bano reports a relationship with GlaxoSmithKline (GSK) that includes: employment and equity or stocks. Samir Diab reports a relationship with GlaxoSmithKline (GSK) that includes: employment. Charalampos Christodoulou reports a relationship with GlaxoSmithKline (GSK) that includes: employment.

Data availability

The data that has been used is confidential.

Acknowledgment

This study was funded by a Digital Design capability project at GlaxoSmithKline (GSK).

Appendix A

A.1 Model-based design of experiments (MBDoE)

MBDoE techniques are used to reduce the parameter uncertainty region with the minimum experimental effort through the optimization of the experiment design vector $\boldsymbol{\varphi}$ (Franceschini and Macchietto, 2008):

$$\boldsymbol{\varphi} = [\boldsymbol{y}_0, \boldsymbol{u}(t), \boldsymbol{t}^{sp}, \tau]^T \tag{A.1}$$

where $\boldsymbol{y}_0$ is the set of initial conditions for the measured variables, $\boldsymbol{t}^{sp}$ is the vector of the output variables sampling times, and τ is the total duration of the experiment.

The optimization problem is conventionally formulated as:

$$\boldsymbol{\varphi}^{opt} = arg\min_{\boldsymbol{\varphi}}\{\psi[\boldsymbol{V}_\theta(\boldsymbol{\theta}, \boldsymbol{\varphi})]\} \tag{A.2}$$

where $\boldsymbol{V}_\theta$ is the the expected variance–covariance matrix of the model parameters θ, while ψ is a selected metric of $\boldsymbol{V}_\theta$ that represents the criterion for the experimental design. In this work, we used the A-optimal criterion, i.e.,$\psi = \boldsymbol{V}_\theta(\theta, \boldsymbol{\varphi})$, which minimizes the dimensions of the enclosing box around the joint confidence region. Optimal values of design variables are obtained through a nonlinear sequential quadratic programming optimization.

Note that different approaches can be used to solve the MBDoE optimization problem, such as worst-case (Nagy and Braatz, 2004) or chance constraint (Ostrovsky et al., 2013) formulations.

$\boldsymbol{V}_\theta([03B8], \boldsymbol{\varphi})$ is evaluated as:

$$\boldsymbol{V}_{\theta}(\boldsymbol{\theta},\boldsymbol{\varphi}) = \left\{ \left(\boldsymbol{V}_{\theta}{}^{0}\right)^{-1} + \sum_{k=1}^{N_{sp}} \sum_{i=1}^{N_y} \sum_{j=1}^{N_y} s_{i,j} \left[\frac{\partial \widehat{y}_i(t_k)}{\partial \theta_l} \frac{\partial \widehat{y}_j(t_k)}{\partial \theta_m} \right]_{l,m=1:1:N_{\theta}} \right\}^{-1}, \tag{A.3}$$

where $s_{i,j}$ is the i,j $^{\text{th}}$ element of the inverse of the measurement error covariance matrix, N_{sp} is the number of sampling intervals, N_y is the number of measured variables, and N_θ is the length of $\boldsymbol{\theta}$. $\boldsymbol{V}_{\theta}{}^{0}$ is the preliminary parameter variance/covariance matrix, which accounts for the initial parameter uncertainty.

A.2 Parameter estimation

Assuming that measurements errors are normally distributed, a maximum likelihood function ϕ^{ML} can be expressed as according to Bard (1974), and model parameters are computed as to maximize its value:

$$\phi^{ML}\left(\boldsymbol{y}, \boldsymbol{\Sigma}_{1,\dots}, \boldsymbol{\Sigma}_N\right) = 2\pi^{\frac{N_y}{2}} \prod_{i=1}^{N} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} \left[(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)^T \boldsymbol{\Sigma}_i{}^{-1} (\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i) \right] \right\} \tag{A.4}$$

where $\boldsymbol{\Sigma}_i$ is the variance–covariance matrix of measurements errors in the i^{th} experiments, with N equal to the total number of samples. $\boldsymbol{y}$ is the vector of measured variables of length N_y, and $\widehat{\boldsymbol{y}}$ is the vector of correspondent model predictions.

The t-value test is used for assessing the precision in parameter θ_i estimation at $(1\ -\alpha)$% confidence level:

$$t_{\theta_i}^{1-\alpha} = \frac{\theta_i}{t(1-\alpha/2, N-N_\theta)\sqrt{v_{ii}}}, \tag{A.5}$$

where $t$at the denominator is the critical value of a t-distribution with $(1-\alpha/2)$ % confidence level and $(N-N_\theta)$ degrees of freedom (N = number of samples; N_θ= number of model parameters); v_{ii} is the ii^{th} term of the variance–covariance matrix. A statistically satisfactory parameter estimation is reached when $t_{\theta_i}^{1-\alpha}$ is greater than the reference t-value t_{ref} with $(1-\alpha)$% confidence level and $(N-N_\theta)$ degrees of freedom

$$t_{ref} = t(1-\alpha, N-N_\theta). \tag{A.6}$$

Confidence intervals (CIs) correspond to the denominator of Eq. (A.5).

In this work, we set the significance level α% equal to 5%.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ces.2023.118972.

References

Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Avraam, M.P., Shah, N., Pantelides, C.C., 1998. Modelling and optimisation of general hybrid systems in the continuous time domain. Comput. Chem. Eng. 22, S221–S228.

Bai, J.P.F., Earp, J.C., Pillai, V.C., 2019. Translational quantitative systems pharmacology in drug development: from current landscape to good practices. AAPS J. 21, 72.

Bano, G., Dhenge, R.M., Diab, S., Goodwin, D.J., Gorringe, L., Ahmed, M., Elkes, R., Zomer, S., 2022. Streamlining the development of an industrial dry granulation process for an immediate release tablet with systems modelling. Chem. Eng. Res. Des. 178, 421–437.

Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press, New York, NY.

Barr, J., Rabitz, H., 2023. Kernel-based global sensitivity analysis obtained from a single data set. Reliab. Eng. Syst. Saf. 235, 109173.

Boukouvala, F., Niotis, V., Ramachandran, R., Muzzio, F.J., Ierapetritou, M.G., 2012. An integrated approach for dynamic flowsheet modeling and sensitivity analysis of a continuous tablet manufacturing process. Comput. Chem. Eng. 42, 30–47.

Box, G.E.P., Hunter, W.G., Hunter, J.S., 1978. Statistics for experimenters. An introduction to design, data analysis and model building. John Wiley and Sons, New York.

Braakman, S., Pathmanathan, P., Moore, H., 2022. Evaluation framework for systems models. CPT Pharmacometr. Syst. Pharmacol. 11 (3), 264–289.

Cenci, F., Bano, G., Christodoulou, C., Vueva, Y., Zomer, S., Barolo, M., Bezzo, F., Facco, P., 2022. Streamlining tablet lubrication design via model-based design of experiments. Int. J. Pharm. 614, 121435.

Chen, Y., Ierapetritou, M., 2020. A framework of hybrid model development with identification of plant-model mismatch. AIChE J. 66, 16996.

Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., Schaibly, J.H., 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. J. Chem. Phys. 59, 3873–3878.

De-Luca, R., Bano, G., Tomba, E., Bezzo, F., Barolo, M., 2020. Accelerating the development and transfer of freeze-drying operations for the manufacturing of biopharmaceuticals by model-based design of experiments. Ind. Eng. Chem. Res 59 (45), 20071–20085.

Destro, F., Barolo, M., 2022. A review on the modernization of pharmaceutical development and manufacturing – trends, perspectives, and the role of mathematical modeling. Int. J. Pharm. 620, 121715.

Destro, F., Nagy, Z.K., Barolo, M., 2022. A benchmark simulator for quality-by-design and quality-by-control studies in continuous pharmaceutical manufacturing - intensified filtration-drying of crystallization slurries. Comp. Chem. Eng. 163, 107809.

Diab, S., Bano, G., Christodoulou, C., Hodnett, N., Benedetti, A., Andersson, M., Zomer, S., 2022a. Application of a system model for continuous manufacturing of an active pharmaceutical ingredient in an industrial environment. J. Pharm. Innov. 17 (4), 1333–1346.

Diab, S., Christodoulou, C., Taylor, G., Rushworth, P., 2022b. Mathematical modeling and optimization to inform impurity control in an industrial active pharmaceutical ingredient manufacturing process. Org. Process Res. Dev. 26 (10), 2864–2881.

Fishman, G.S., 1995. Monte Carlo: Concepts, algorithms, and applications. Springer-Verlag, N.Y., USA.

Franceschini, G., Macchietto, S., 2008. Model-based design of experiments for parameters precision: state of the art. Chem. Eng. Sci. 63, 4864–4872.

Geladi, P., Kowalski, B.R., 1996. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17.

Gunther, J.C., Baclaski, J., Seborg, D.E., Conner, J.S., 2009. Pattern matching in batch bioprocesses – comparisons across multiple products and operating conditions. Comput. Chem. Eng. 33 (1), 88–96.

Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. Reliab. Eng. Syst. Saf. 52 (1), 1–17.

Hotelling, H., 1993. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24, 417–441.

Iman, R.L., Helton, J.C., 1988. An investigation of uncertainty and sensitivity analysis techniques for computer models. Risk Anal. 8 (1), 71–90.

Jaeckle, C.M., Macgregor, J.F., 1998. Product design through multivariate statistical analysis of process data. AIChE J. 44 (5), 1105–1118.

Jaeckle, C.M., MacGregor, J.F., 2000. Industrial applications of product design through the inversion of latent variable models. Chemom. Intell. Lab Syst. 50 (2), 199–210.

Kaiser, H.F., 1991. Coefficient alpha for a principal component and the Kaiser-Guttman rule. Psychol. Rep. 68 (3), 855–858.

Kawakita, K., Lüdde, K.-H., 1971. Some considerations on powder compression equations. Powder Technol. 4 (2), 61–68.

Kucherenko, S., Albrecht, D., Saltelli, A., 2015. Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques. ArXiv150502350 Stat.
Kuentz, M., Leuenberger, H., 1998. Modified Young's modulus of microcrystalline cellulose tablets and the directed continuum percolation model. Pharm. Dev. Technol. 3 (1), 13–19.
Kushner, J., 2012. Incorporating Turbula mixers into a blending scale-up model for evaluating the effect of magnesium stearate on tablet tensile strength and bulk specific volume. Int. J. Pharm. 399, 19–30.
Ljung, L., 1986. System Identification: Theory for the User. Prentice-Hall Inc, USA.
MacGregor, J.F., Bruwer, M.-J., 2008. A framework for the development of design and control spaces. J. Pharm. Innov. 3 (1), 15–22.
Mahalanobis, P.C., 1930. On tests and measures of groups divergence. J. Asiat. Soc. Bengal 26, 541–588.
Markl, D., Yassin, S., Wilson, D.I., Goodwin, D.J., Anderson, A., Zeitler, J.A., 2017. Mathematical modelling of liquid transport in swelling pharmaceutical immediate release tablets. Int. J. Pharm. 526 (1-2), 1–10.
McLean, K.A.P., McAuley, K.B., 2012. Mathematical modelling of chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. Can. J. Chem. Eng. 90 (2), 351–366.
Metta, N., Ghijs, M., Schäfer, E., Kumar, A., Cappuyns, P., Van Assche, I., Singh, R., Ramachandran, R., De Beer, T., Ierapetritou, M., Nopens, I., 2019. Dynamic flowsheet model development and sensitivity analysis of a continuous pharmaceutical tablet manufacturing process using the wet granulation route. Processes 7, 234.
Miao, H., Xia, X., Perelson, A.S., Wu, H., 2011. On identifiability of nonlinear ODE models and applications in viral dynamics. SIAM Rev. 53 (1), 3–39.
Montes, F.C.C., Gernaey, K., Sin, G., 2018. Dynamic plantwide modeling, uncertainty, and sensitivity analysis of a pharmaceutical upstream synthesis: ibuprofen case study. Ind. Eng. Chem. Res. 57 (30), 10026–10037.
Montgomery, D.C., 2013. Introduction to Statistical Quality Control. Wiley, Hoboken, NJ.
Moreno-Benito, M., Lee, K.T., Kaydanov, D., Verrier, H.M., Blackwood, D.O., Doshi, P., 2022. Digital twin of a continuous direct compression line for drug product and process design using a hybrid flowsheet modelling approach. Int. J. Pharm. 628, 122336.
Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161–174.
Nagy, Z.K., Braatz, R.D., 2004. Open-loop and closed-loop robust optimal control of batch processes using distributional and worst-case analysis. J. Process Control 14 (4), 411–422.
Nassar, J., Williams, B., Davies, C., Lief, K., Elkes, R., 2021. Lubrication empirical model to predict tensile strength of directly compressed powder blends. Int. J. Pharm. 592, 119980.
Ostrovsky, G.M., Ziyatdinov, N.N., Lapteva, T.V., 2013. Optimal design of chemical processes with chance constraints. Comput. Chem. Eng. 59, 74–88.
Peppas, N., Colombo, P., 1989. Development of disintegration forces during water penetration in porous pharmaceutical systems. J. Control. Release 10 (3), 245–250.
Pitt, K.G., Heasley, M.G., 2013. Determination of the tensile strength of elongated tablets. Powder Technol. 238, 169–175.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S. (Eds.), 2007. Global Sensitivity Analysis. The Primer. Wiley.
Schenkendorf, R., Xie, X., Rehbein, M., Scholl, S., Krewer, U., 2018. The impact of global sensitivities and design measures in model-based optimal experimental design. Processes 6, 27.
Shahmohammadi, A., McAuley, K.B., 2020. Using prior parameter knowledge in model-based design of experiments for pharmaceutical production. AIChE J. 66, 17021.
Sobol, I.M., 1993. Sensitivity estimates for nonlinear mathematical models. Math. Model. Comput. Exp. 407–414.
Song, E., Nelson, B.L., Staum, J., 2016. Shapley effects for global sensitivity analysis: theory and computation. Soc. Ind. Appl. Math. 4 (1), 1060–1083.
Sorenson, H.W., 1980. Parameter estimation: Principles and problems. Marcel Dekker, New York, NY.
Tian, H., Bhalode, P., Razavi, S.M., Koolivand, A., Muzzio, F.J., Ierapetritou, M.G., 2022. Characterization and propagation of RTD uncertainty for continuous powder blending processes. Int. J. Pharm. 628, 122326.
The United States Pharmacopeial Convention, 2011.
White, L.R., Molloy, M., Shaw, R.J., Reynolds, G.K., 2022. System model driven selection of robust tablet manufacturing processes based on drug loading and formulation physical attributes. Eur. J. Pharm. Sci. 172, 106140.
Wilson, D., Wren, S., Reynolds, G., 2012. Linking dissolution to disintegration in immediate release tablets using image analysis and a population balance modelling approach. Pharm. Res. 29 (1), 198–208.
Wise, B.M., Gallagher, N.B., 2006. The process chemometrics approach to process monitoring and fault detection. J. Chemometr. 12, 301–321.
Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method. In Matrix Pencils 286–293.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58 (2), 109–130.
Xie, X., Schenkendorf, R., Krewer, U., 2019. Efficient sensitivity analysis and interpretation of parameter correlations in chemical engineering. Reliab. Eng. Syst. Saf. 187, 159–173.
Yang, O., Tao, Y., Qadan, M., Ierapetritou, M., 2022. Process design and comparison for batch and continuous manufacturing of recombinant adeno-associated virus. J. Pharm. Innov. 49, 2215–2220.
Zineh, I., 2019. Quantitative systems pharmacology: a regulatory perspective on translation. CPT Pharmacometr. Syst. Pharmacol. 8 (6), 336–339.
Zwick, W.R., Velicer, W.F., 1986. Comparison of five rules for determining the number of components to retain. Psychol. Bull. 99, 432–442.